Insect Biochemistry and Molecular Biology

A phylogenomics approach to characterizing sensory neuron membrane proteins (SNMPs) in Lepidoptera

--Manuscript Draft--

Manuscript Number:	IB_2019_232R2
Article Type:	Full Length Article
Keywords:	SNMP; Sensory neuron membrane proteins; olfactory receptor; Insect olfactory system; Bombyx mori
Corresponding Author:	Wei Xu Murdoch University Murdoch, Western Australia Australia
First Author:	Hui-Jie Zhang
Order of Authors:	Hui-Jie Zhang
	Wei Xu
	Quan-mei Chen
	Lena Sun
	Alisha Anderson
	Qingyou Xia
	Alexie Papanicolaou
Abstract:	Sensory neuron membrane proteins (SNMPs) play a critical role in the insect olfactory system but there is a deficit of functional studies beyond Drosophila . Here, we use a combination of available genome sequences, manual curation, genome and transcriptome data, phylogenetics, expression profiling and gene knockdown to investigate SNMP superfamily in various insect species with a focus on Lepidoptera. We curated 81 genes from 36 insect species and identified a novel lepidopteran SNMP gene family, SNMP3. Phylogenetic analysis shows that lepidopteran SNMP3, but not the previously annotated lepidopteran SNMP2, is the true homologue of the dipteran SNMP2. Digital expression, microarray and qPCR analyses show that the lepidopteran SNMP1 is specifically expressed in adult antennae. SNMP2 is widely expressed in multiple tissues while SNMP3 is specifically expressed in the larval midgut. Microarray analysis suggest SNMP3 may be involved in the silkworm immunity response to virus and bacterial infections. We functionally characterized SNMP1 in the silkworm using RNA interference (RNAi) and behavioral assays. Our results suggested that Bombyx mori SNMP1 is a functional orthologue of the Drosophila melanogaster SNMP1 and plays a critical role in pheromone detection. Split-ubiquitin yeast hybridization study shows that BmorSNMP1 has a protein-protein interaction with the pheromone receptor (BmorOR1), and the co-receptor (BmorOrco). Concluding, we propose a novel molecular model in which BmorOrco, BmorSNMP1 and BmorOR1 form a heteromer in the detection of the silkworm sex pheromone bombykol.
Response to Reviewers:	

The reviewers recommend reconsideration of your manuscript following major revisions and modification. Both reviewers were not pleased that all the issues raised in the previous review were not addressed. It is critical you do that. Hence I invite you to resubmit your manuscript after addressing ALL the comments below.

Thanks for your suggestions! We ve revised the manuscript and address the changes in yellow according to the reviewers comments.

When revising your manuscript, please consider all issues mentioned in the reviewers' comments carefully: please outline every change made in response to their comments and provide suitable rebuttals for any comments not addressed. Please note that your revised submission may need to be re-

Thanks for your suggestions! We ve revised the manuscript and address the changes in yellow according to the reviewers comments. And we have written rebuttals for the comments.

Reviewer #1: This reviewer has impression on this manuscript because experimental data has still somewhat questionable. Authors should re-analyze the data in detail.
"100 insects were used totally for each group. That means each individual one is a biological repeat in each group and two experimental repeats."
We rewrote the sentence in 2.5 RNAi bioassay.
100 male pupae were used and each one is an individual biological replicate. All 100 pupae were kept at 25 °C until eclosion.

50 individuals were used for qRT-PCR experiment? The number of repeats per experiment of fig2、fig3 should be described.

In Material and method part. 2.1. Insect materials We added the how the tissues were collected for the RT-PCR and qRT-PCR.
Multiple tissues studied for RT-PCR and real-time PCR are listed as follows: 100 antennae, 50 mouthparts and five midguts were dissected from larvae at 3rd day in 5th instar; 20 adult antennae, 10 heads without antennae, five midguts, 20 legs and 10 wings were dissected from emerging adults. All collected tissues were immediately immersed into liquid nitrogen and stored at -80 °C until use. *Helicoverpa armigera* and *D. melanogaster* (Canton-S) were provided by CSIRO Ecosystem Sciences. Same tissues of *H. armigera* were collected at the similar age by using the same methods as those for *B. mori*. Larval heads and midguts of *D. melanogaster* were dissected from 50 3rd instar larvae and all the other adult tissues were collected 50 4-day aged virgin adult flies.

In result parts, we all described how many replicates were prepared.

Three repeats were performed in real-time (qRT-PCR).

In Figure 2F-G, the micro-array data was built by Cheng et al., 2016. There are details for samples collection.

For Figure 3, we mentioned 100 male pupae were used and each one is an individual biological replicate. All pupae were kept at 25 °C until eclosion.

In methods 2.5 the last paragraph:

After behavioural testing, 80 antennae were collected from the control and ddH2O moths respectively. The sample from RNAi treated groups were collected from all male moth which didn't find female in 20min and five from those after 15 min randomly. Totally 43 male moths from BmorSNMP1^{-}/BmorOR1^{+} group, 42 from BmorSNMP1^{+}/BmorOR1^{-} group and 39 from BmorSNMP1^{-}/BmorOR1^{-} group were used to collect antennae for qRT-PCR analysis. (Supplementary Table 5)

In results 3.5 the beginning of the last paragraph: After behavioral testing, the expression level of BmorSNMP1 and BmorOR1 .

"In figure 2C-E, it could show mean+-sem, and degrees of freedom for three repeats. One-way ANOVA has been used and followed by Turkey's multiple comparisons test." authors must use at least four repeats because of generating normal distribution.

Here we used three replicates following the literatures below who used one-way ANOVA Turkey-test for qRT-PCR (n=3) according to the reference below

Melatonin Pretreated Blastocysts along with Calcitonin Administration Improved Implantation by Upregulation of Heparin Binding-Epidermal Growth Factor Expression in Murine Endometrium. Moghani-Ghoroghi F, Moshkdanian G, Sehat M, Nematollahi-Mahani SN, Ragerdi-Kashani I, Pasbakhsh P. Cell J. 2018 Jan;19(4):599-606.

Reviewer #2: This is my second review of this manuscript, so I will keep this brief and address the authors' responses to my previous comments. Great paper!
I. Major Revisions.
1. Phylogeny of SNMPs. My initial concerns here are resolved. However, seeing the node support and topology in the new trees, I am not sure it is necessary to include Fig. 1A. Many nodes in Fig. 1B have high confidence values but lose resolution in Fig. 1A (likely because the divergent CD36 genes interfere with the MSA). Moreover, Fig. 1A is used only to demonstrate that the SNMPs form a monophyletic clade (L344 in the revised ms), but this is already understood (e.g., Vogt et al. 2009. IBMB 39: 448-456). Thus, I think 1A could be removed at no loss to the ms, with perhaps instead a few representative CD36 genes used to root Fig. 1B.

Figure 1A was deleted.

In 2.2 Gene identification and phylogenetic analysis:
family has been deleted from the sentence Reported sequences of SNMPs and CD36 family members from Lepidoptera, Diptera, Coleoptera and Hymenoptera were downloaded from NCBI or relevant sources (Forstner et al., 2008; Gu et al., 2013; Li and Qin, 2011; Liu et al., 2013a; Liu et al., 2013b; Liu et al., 2012; Nichols and Vogt, 2008; Rogers et al., 2001a; Rogers et al., 2001b; Rogers et al., 1997; Vogt et al., 2009). ”

In 3.2 Phylogenetic characterization:
has been replaced by (Vogt et al., 2019) in the sentence

Figure 1B was changed as Figure 1 in text.

In Figure 1 legends:
Those sentences have been changed as below: Figure 1. Phylogenetic analysis of insect SNMP homologs. ML tree has been built by IQ-TREE. Blue, green and purple branches present SNMP1, SNMP2 and SNMP3. The bootstrap is 1000 replications .

2. Incidentally, the present tree in Fig. 1B indicates a root, but no outgroup was given. If this was an unrooted tree, the central line should be removed.

The central line is shown ancestral lineage rather than tip. We deleted this line from the tree.

3. Presentation of the research. My concerns here are mostly resolved, but: Figure 5 is now included in the ms, but still lacks a figure legend.
We added the legend.
Figure 5. The new mode of Orco, SNMP1 and pheromone receptor function mechanism. In *B. mori*, sex pheromone compound (bombykol) was transported by pheromone binding protein (PBP) to the functional complex which is composed by pheromone receptor (BmorOR1), co-receptor (BmorOrco) and BmorSNMP1, to activate the moth responses.

II. Minor Revisions. My concerns here are mostly resolved, with one note below: Table 1. DponSNMP1/1a/2 were removed from Table 1 (but are still present in the phylogeny). My initial comment regarding these genes was only to correct their names in the original table - I have no other concern with including them, and apologize if I gave that impression!

In the table 1, DponSNMP1/1a/2 sequences were added back.

Highlights:

1. Lepidopteran SNMP3, but not the SNMP2, is the true homologue of the dipteran SNMP2.

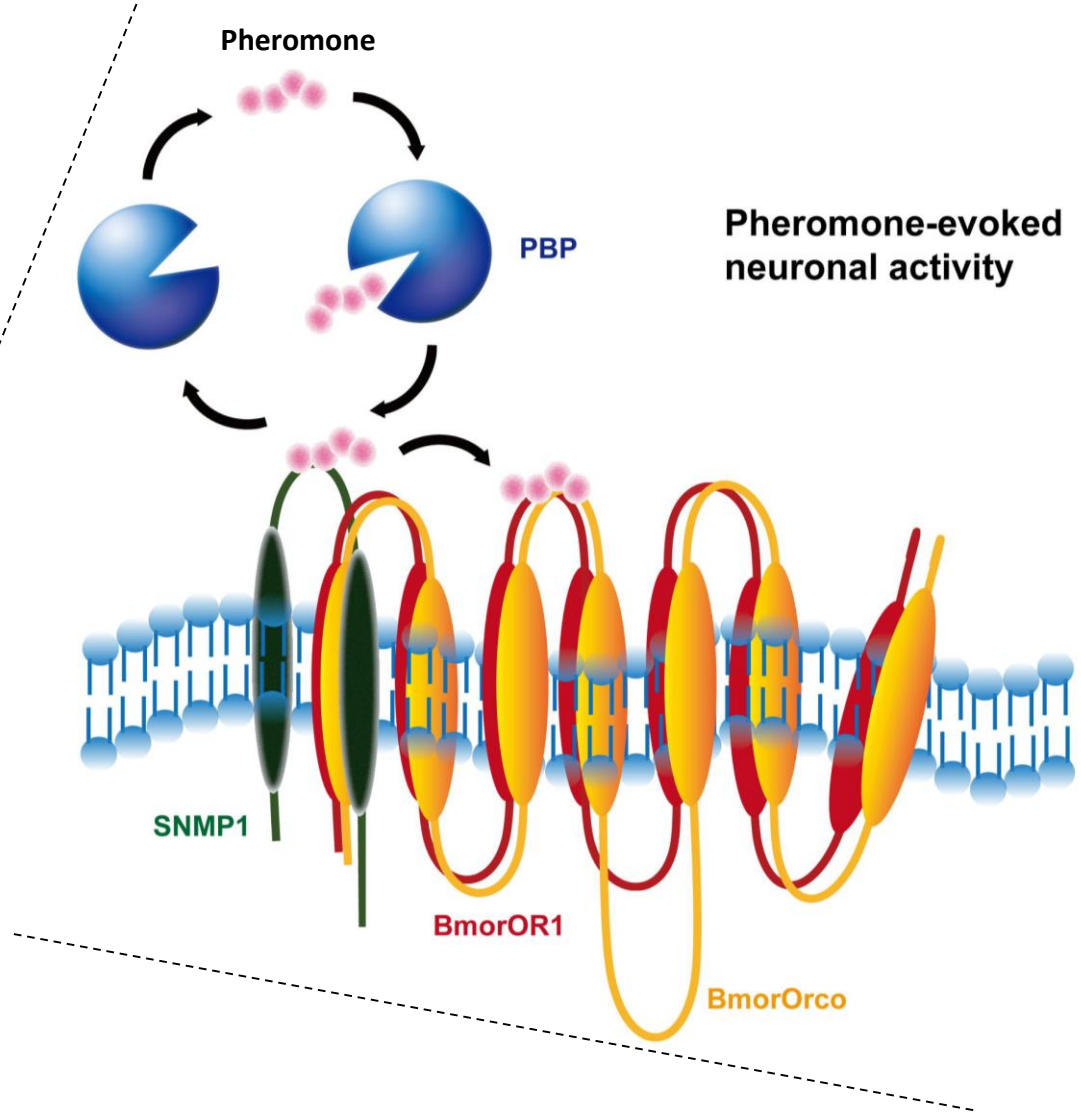
2. Lepidopteran SNMP1 is specifically expressed in adult antennae. SNMP2 is widely expressed while SNMP3 is specifically expressed in the larval midgut.

3. *Bombyx mori* SNMP1 is a functional orthologue of the *Drosophila* SNMP1 and plays a critical role in the pheromone detection.

4. BmorSNMP1 has a protein-protein interaction with the pheromone receptor (BmorOR1), and the co-receptor (BmorOrco).

5. A novel molecular model was proposed: BmorOrco, BmorSNMP1 and BmorOR1 form a heteromer in the detection of the sex pheromone.

Graphical Abstract



A phylogenomics approach to characterizing sensory neuron membrane proteins (SNMPs) in Lepidoptera

Hui-Jie Zhang[1,2,3,#], Wei Xu[4], Quan-mei Chen[1], Le-Na Sun[1], Alisha Anderson[2], Qing-You Xia[1,*], Alexie Papanicolaou[3,5,*]

[1]*State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400715, China*

[2]*CSIRO Food Futures Flagship, Canberra, ACT 2601, Australia*

[3]*CSIRO Ecosystem Sciences, Canberra, ACT 2601, Australia*

[4]*Agricultural Sciences, Murdoch University, Murdoch, WA 6150, Australia*

[5]*Hawkesbury Institute for the Environment, Western Sydney University, Richmond 2753, Australia*

[#]*Current affiliation is Key Laboratory of Biochemistry and Molecular Pharmacology of Chongqing, Chongqing Medical University, Chongqing, People's Republic of China, 400016*

*Corresponding authors:

Qing-You Xia, Phone: +86 23 68250099, Fax: +86 23 68251128; Email: xiaqy@swu.edu.cn

Alexie Papanicolaou, Phone: +61 2 4570 1385, Email: A.Papanicolaou@westernsydney.edu.au

Abstract: Sensory neuron membrane proteins (SNMPs) play a critical role in the insect olfactory system but there is a deficit of functional studies beyond *Drosophila*. Here, we use a combination of available genome sequences, manual curation, genome and transcriptome data, phylogenetics, expression profiling and gene knockdown to investigate SNMP superfamily in various insect species with a focus on Lepidoptera. We curated 81 genes from 36 insect species and identified a novel lepidopteran SNMP gene family, SNMP3. Phylogenetic analysis shows that lepidopteran SNMP3, but not the previously annotated lepidopteran SNMP2, is the true homologue of the dipteran SNMP2. Digital expression, microarray and qPCR analyses show that the lepidopteran SNMP1 is specifically expressed in adult antennae. SNMP2 is widely expressed in multiple tissues while SNMP3 is specifically expressed in the larval midgut. Microarray analysis suggest SNMP3 may be involved in the silkworm immunity response to virus and bacterial infections. We functionally characterized SNMP1 in the silkworm using RNA interference (RNAi) and behavioral assays. Our results suggested that *Bombyx mori* SNMP1 is a functional orthologue of the *Drosophila melanogaster* SNMP1 and plays a critical role in pheromone detection. Split-ubiquitin yeast hybridization study shows that BmorSNMP1 has a protein-protein interaction with the pheromone receptor (BmorOR1), and the co-receptor (BmorOrco). Concluding, we propose a novel molecular model in which BmorOrco, BmorSNMP1 and BmorOR1 form a heteromer in the detection of the silkworm sex pheromone bombykol.

Keywords: Sensory neuron membrane proteins, SNMP, Olfactory receptor, Insect olfactory system, *Bombyx mori*

1. Introduction

With the increasing throughput of draft genome sequencing we are now able to complete a genome project within months. These genome projects are however working drafts and may harbor misassemblies or misannotations, requiring substantial curation before biological inferences can be made. To date there has been limited appreciation for professional curation by the wider genomic community. By curation here, we not only refer to a manual edition of the underlying gene model (i.e. structural annotation) but also functional assignment. This is an important activity: Any automated assignment of orthology – such as one based solely on the information contained within the protein sequences – does not rely upon, or imply, an orthology of function. Further, with incomplete ascertainment or without a robust phylogenetic framework, an assignment of orthology by descent may also be erroneous.

It is generally a challenge to identify accurately structural or functional annotations for diverged genes. An additional tool that underpins our ability to produce higher quality of both structural and functional annotations is expression profiling. Using RNASeq (i.e. high throughput sequencing) we can accurately detect both the open reading frames and untranslated regions of transcripts. In addition, expression level profiling RNASeq and quantitative PCR (qPCR) can provide strong evidence for functional annotation assignment. Indeed, together with a knock-down/out technique such as RNA interference (RNAi) or clustered regularly interspaced short palindromic repeats (CRISPR), these approaches give us one of the strongest lines of evidence what the function of a gene may be. Having accurate assignments is important to the wider biology community: The traditional concept of nomenclature implies that genes between two species that have the same name are orthologous either by function or descent.

In this work we use a combination of available genome sequences, manual curation, genome and transcriptome data, phylogenetics, expression profiling and gene knockdown to investigate a family of genes from the CD36 superfamily in insects. The CD36 superfamily codes for membrane-bound scavenger proteins with a variety of ligands and signalling functions. In insects, the best characterized of these functions relate to signalling in response to environmental cues, such as diet, stress, innate

immunization and chemoreception. There are a total of three CD36 families and here we focus on the sensory neuron membrane proteins (SNMPs) (Vogt et al., 2009). The first SNMPs were identified from lepidoptrea (Rogers et al., 2001a; Rogers et al., 2001b; Rogers et al., 1997). SNMP1 expression was shown to be highly expressed in pheromone receptor neurons of trichoid sensilla (Rogers et al., 1997). Further evidence indicates that SNMP1 in *Drosophila melanogaster* is a cofactor along with DmelOR67d/Orco of the sex pheromone detection system and, therefore, critical for species-recognition (Benton et al., 2007). In addition, it was recently shown that SNMP1 interacts directly with DmelOR22a, rather than Orco using Förster resonance energy transfer (FRET) system (German et al., 2013). The SNMP2 protein has yet to be fully characterized, but moth SNMP2 is known to express in supporting cells around odorant sensitive neurons (OSNs) in olfactory antennae sensilla (Forstner et al., 2008; Rogers et al., 2001b). Recent studies also showed that the moth SNMP2 is broadly and highly expressed in antennae, legs and wings (Liu et al., 2014; Zhang et al., 2015; Zhang et al., 2013). However, *Drosophila* transcriptome profiling from Flybase has shown SNMP2 has limited expression in tracheal, reproductive and integumentary systems (http://flybase.org/reports/FBgn0035815.html).

In our present work, the objective was to characterise the SNMP family in moths, but an advantage is that we began with a phylogenomic approach to determine the true orthology of SNMP genes. Therefore, we first fully ascertained the SNMP family and generated a robust and accurate phylogenomic model. As a result, we discovered that a novel SNMP gene arose in Lepidoptera and named SNMP3. We then proceeded to integrate public transcriptomic evidence to build hypotheses around function. Second, by integrating the expression data with the classic phylogenetic clustering, we showed that the family has been misannotated, likely due to previous dependence on BLAST similarity searches rather than a phylogenomic approach. Third, we provided multiple lines of evidences including RNAi, split-ubiquitin yeast hybridization system and expression data, to show that - despite its sequence divergence, the function of the SNMP1 complex is conserved between Diptera and Lepidoptera, and is essential to mating recognition. We recommend it as a model for studying protein co-evolution driven by assortative mating.

2. Material and Methods

2.1 Insect materials

B. mori, Dazao strain, larvae and adults were reared on mulberry leaves at 25°C by the State Key Laboratory of Silkworm Genome Biology, Southwest University. Multiple tissues studied for RT-PCR and real-time PCR are listed as follows: 100 antennae, 50 mouthparts and five midguts were dissected from larvae at 3rd day in 5th instar; 20 adult antennae, 10 heads without antennae, five midguts, 20 legs and 10 wings were dissected from emerging adults. All collected tissues were immediately immersed into liquid nitrogen and stored at -80 °C until use. *Helicoverpa armigera* and *D. melanogaster* (Canton-S) were provided by CSIRO Ecosystem Sciences. Same tissues of *H. armigera* were collected at the similar age by using the same methods as those for *B. mori*. Larval heads and midguts of *D. melanogaster* were dissected from 50 3rd instar larvae and all the other adult tissues were collected 50 4-day aged virgin adult flies.

2.2 Gene identification and phylogenetic analysis

Reported sequences of SNMPs members from Lepidoptera, Diptera, Coleoptera and Hymenoptera were downloaded from NCBI or relevant sources (Forstner et al., 2008; Gu et al., 2013; Li and Qin, 2011; Liu et al., 2013a; Liu et al., 2013b; Liu et al., 2012; Nichols and Vogt, 2008; Rogers et al., 2001a; Rogers et al., 2001b; Rogers et al., 1997; Vogt et al., 2009). The sequences of SNMP in *D. melanogaster*, *Anopheles gambiae*, *Apis mellifera* and *Tribolium castaneum* used here are identical to those used previously (Nichols and Vogt, 2008). We searched known insect genome and/or transcriptome (Table 1) using tBLASTn (Altschul et al., 1997). Candidate genes were annotated manually in Geneious R7 created by Biomatters (http://www.geneious.com/). Multiple transmembrane domains of receptors were predicted by TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/). Amino acid sequences were aligned with MUSCLE (Edgar, 2004) using default parameters. The maximum likelihood (ML) reconstructions were performed using IQ-TREE v1.6.3 (Nguyen et al., 2015) with 1,000 ultrafast bootstrap (Hoang et al., 2018) and 1,000 SH-aLRT replicates (Guindon et al.,

2010) estimated. Substitution models were estimated with ModelFinder (Kalyaanamoorthy et al., 2017) built-in in IQ-TREE (Figure 1).

2.3 RT-PCR

Total RNA was extracted by Trizol (Promega, USA) according to the protocol provided by the manufacturer. Total RNA was treated using DNase I, then quantified and qualified by NanoDrop ND-2000 (Thermo Scientific, USA). First strand cDNA was synthesized using SuperScript III First-Strand Synthesis Super Mix (Invitrogen, USA). Expressions of SNMPs in multiple tissues and during development were tested by semi-qRT-PCR. All primers for both semi-qRT-PCR and qRT-PCR are listed in Supplementary Table 1. BmorRPL3 was used as a reference gene in qRT-PCR because its expression has been detected stable in various stages and tissues of *B. mori*. Three experimental repeats have been performed for both semi-qRT-PCR and qRT-PCR.

2.4 In-silico expression analysis

To gain insight into the function of SNMPs involving in immunization, microarray-annotated silkworm datasets (http://www.silkdb.org/silkdb) were screened with tBLASTn (Altschul et al., 1997). The regulated expression of genes in individual larva are denoted by ratio comparing the gene expressed level in intact larva (non-induced control). For those non-induced control, the same volume of ddH_2O was mixed in the feed for the silkworm and the rearing conditions kept the same as bacteria or virus induced (Cheng et al., 2016).

2.5 RNAi bioassay

RNAi was performed using dsRNA. The position and the length of dsRNA are specifically designed according to the topology and functional domain of each gene, which covered the main functional domains for specific targeting. The dsRNA of BmorSNMP1 is spanning the second transmembrane domain (309-520 AA) (Fwd 5'

GCTAATACGACTCACTATAGGGAGATACAATGGGATTAAGACGA 3'; Rev 5' GCTAATACGACTCACTATAGGGAGATTTGGCTGGTTCTTGATT 3'), while that for BmorOR1 is spanning from the third to fifth transmembrane domains (109-255 AA) (Fwd 5' GCTAATACGACTCACTATAGGGAGATCTTGTATTAACTGGTCGCTTCA 3'; Rev 5' GCTAATACGACTCACTATAGGGAGATGGCTGGCTTTAGGTCTCG). All primers are fused by T7 promoter. dsRNA was prepared using RiboMAX Large Scale System-T7 kit (Promega, UAS) according to manufacturer's instructions. 30 μg dsRNA in 10 μl ddH_2O was micro-injected into one antenna on the 4th day of pupae. Various concentrations of dsRNA have been tested in preliminary experiments while 30μg dsRNA in 10 μl ddH_2O yielded the low rates for non-eclosion pupae and high numbers of adults which behaved differently from control groups. Intact and ddH_2O injected pupae are used as control. Chlortetracycin Hydrochloride Eye Ointment was used to prevent bacterial contamination after injection. 100 male pupae were used and each one is an individual biological replicate. All pupae were kept at 25 °C until eclosion.

Within 48 hr after eclosion, virgin male moths were tested. One male was placed into a glass cultivate vessel (internal diameter = 180 mm) first and then one virgin female moth was added at the longest distance away from the male within the glass cultivate vessel. The movements of all male moths were recorded for 20 min and traces were analyzed by Tracker (http://www.cabrillo.edu/~dbrown/tracker).

After behavioural testing, 50 antennae were collected from the moths. The sample from RNAi treated groups were collected from all male moth which didn't find female in 20 min and five from those after 15 min randomly. Totally 43 male moths from BmorSNMP1$^-$/BmorOR1$^+$ group, 42 from BmorSNMP1$^+$/BmorOR1$^-$ group and 39 from BmorSNMP1$^-$/BmorOR1$^-$ group were used to collect antennae for qRT-PCR analysis.. The relative expression of BmorSNMP1 and BmorOR1 in male antennae after RNAi was analysed by qRT-PCR using SYBR Premix Ex Taq^{TM} (Perfect Real Time, Takara) according to the instructions manual for ABI 7500 fast real-time PCR system. Three experimental replicates were performed. The relative gene expression data were analyzed using the method described previously (Livak and Schmittgen, 2001).

2.6 Split-ubiquitin yeast hybridization

The yeast two-hybrid system is designed based upon the principle of protein fragment complementation, which is the most widely used and powerful method to identify and characterize novel protein interactions. The split-ubiquitin yeast two-hybridization system is performed to discover the novel interaction between two transmembrane proteins (Fetchko and Stagljar, 2004; Scheper et al., 2003; Stagljar et al., 1998; Thaminy et al., 2003). Over the past decade, the system has been adapted to cover an increasingly wide range of applications (Snider et al., 2010). In this study, it was performed using DUALmembrane pairwise interaction kit according to the manual (Dualsystems Biotech, Switzerland) and with the method used (Lentze and Auerbach, 2008). Primers and constructs for fusion protein expression constructs were listed in Supplementary Table 2. Vectors were chosen to adapt for the transmembrane topology of proteins. To test for correct expression and functionality of the bait protein in yeast, pOST1-NubI were used as positive control prey plasmid leading to the expression of a fusion membrane protein with NubI residented in ER. pPR3N prey vector providing NubG and pBT3STE and pBT3SUC bait vector providing Cub were used in purpose. SNMP1 and BmorOR2 genes were cloned into pPR3N prey, while cis-vaccinyl acetate receptor DmelOR22a and pheromone receptors (BmorOR1 and DmelOR67d and HarmOR13) were cloned into pBT3STE and pBT3SUC bait vectors separately. Saccharomyces cerevisiae yeast (NMY32) were suspended amplified in YPDA (1% yeast extract, 2% Trytone, 2% Glucose and 0.02% Adenine) at 30ºC and 225rpm until OD600>1.5. Binary constructs were subsequently co-transformed into yeast which were then cultured in SD-Trp-Leu (SD-T/L) and SD-Trp-Leu-His-Ade (SD-T/L/H/A) defective selection solid medium. The expression of three reporter genes of HIS3, ADE2 and LacZ in yeast could be induced due to the interaction between two fusion expressed proteins. The results are represented as positive clones in selection medium. The activity of β-galactosidase encoded by LacZ gene in yeast could be stained by X-gal solution at room temperature and then detected by microplate reader. Data were analyzed with GraphPad Prism 5 using the two-tailed Student’s t-test for comparisons between the potential pairwise interaction and negative control (pPR3N / pBT3STE).

2.7 Statistic analysis

All data was shown as mean ± SEM. The results were graphed and analyzed in GraphPad Prism 5.0 (GraphPad, USA). Statistically significant differences ($\alpha<0.05$) among various groups were measured with ANOVA analysis followed by Tukey's multiple comparisons test (Moghani-Ghoroghi et al., 2018). Survival curves with log-rank test were used to compare the behavioural data for each male approaching female.

3. Results

3.1 Full ascertainment of SNMPs in various insect species

Using official genome annotations, community contributions and similarity searches we collected a total of 81 SNMP genes from 36 insect species as distributed amongst seven orders (Table 1). To fully ascertain the gene family, we manually curated each species for which we had transcriptome or genome assemblies available (Supplementary Table 3), creating a high confidence dataset of 19 species from all seven orders. We also searched two non-insect invertebrate species (*Daphnia pulex* and *Caenorhabditis elegans*) but we did not find any SNMP genes that could be unambiguously inferred as orthologous to our insect core set. Of these 81 SNMP genes, 56 had accurate public gene structures; we edited two structures, contributed a further 16 new annotations and report four partial genes found in public transcriptome databases (Table 1). Overall, 19 of the species had full ascertainment for the SNMP gene family (i.e. the correct public annotations for all SNMPs orthologues known for each order) and 16 had some missing or incorrect annotations (Table 1). It is notable that in spite of available full genome sequence, we did not identify the homologue of *Apis mellifera* SNMP2 in *Nasonia vitripennis* and *Solenopsis invicta*. In addition, *N. vitripennis* had four tandemly duplicated *SNMP1* genes (XP_001606602, XP_001606682, XP_001606675, XP_001606692), which we could not verify due to the lack of primary data for the assembly. Further, the pea aphid (*Acyrthosisphon pisum*) had only one SNMP gene family member, even though another Hemipteron with a full genome sequence (bedbug, *Cimex lectularius*) had three gene family members. Finally, we annotated a new insect order, Psocodea, using the published genome sequence of body louse (*Pediculus humanus corporis*) (Kirkness et al., 2010), identifying two SNMP genes. Due to the lack of RNAseq or comparative data, however, we were not able to characterize the full length of these genes.

Using the *D. melanogaster* homologue from FlyBase (FBgn0035815) we identified the *D. pseudoobscura* orthologue via OrthoDB (cluster EOG79D9HT) and named *DpseSNMP2* (FBgn0080333; XP_002024979), which had two extraneous, unsupported exons (4^{th} and 5^{th}). For *Apis mellifera* (Hymenoptera) we found that the *A. mellifera*

RefSeq annotation of *SNMP2* (XP_001121085) was incorrect. We then built a new gene based on GB12830-PA by identifying the missing N-terminus and a middle exon. For *S. invicta* (Hymenoptera) we used the genome raw data (GCA_000188075.1) to annotate and manually extend two hypothetical proteins (EFZ18285 and EFZ18363), which are tandem arrayed in scaffold_02797 (Table 1).

Further, we identified new genes in Psocodea, Lepidoptera, Diptera and Hemiptera. For *P. humanus* (Psocodea) we found two SNMP genes from the publicly genome sequence available from Ensembl (scaffolds DS235882 and DS235862). Without access to RNAseq data, both gene annotations are missing their C-termini and SNMP1 is missing the N-terminus.

For *C. lectularius* (Hemiptera) three SNMPs were identified using a genome assembly (GCA_001460545.1; https://apollo.nal.usda.gov/cimlec/sequences). The full-length sequence from SNMP1 (512 amino acid) was annotated in scaffold 41 and verified with an RNAseq assembly. A new assembly (GCA_000648675.3) was produced with a more correct gene model (XP_014251146). For SNMP2a, a partial gene encoding 448 amino acids was identified from scaffold_56 and the DNA- and RNAseq data point towards a misassembly for the last exon. Finally, a full-length SNMP2b encoding 510 amino acid was curated from scaffold_1 and verified with an RNAseq assembly.

For *A. pisum* (Hemiptera): using the second genome version (GCA_000142985.2) and we identified the complete CDS of SNMP1 from scaffold EQ121771. However, we did not identify any sequences homologous to the other SNMPs.

For *Glossina morsitans* (Diptera): using the genome (https://pre.vectorbase.org/) a partial SNMP1 and a full ORF of SNMP2 orthologues were identified from scaffolds scf7180000644980 and scf7180000648879 respectively. Transcriptome data was available via VectorBase's Sanger EST transcripts, but we found support only for the last exon of SNMP2 (FM961280.1).

For *B. mori* (Lepidoptera) we found a novel gene in scaffold DF090341.1, which was named *SNMP3*. Two automated predictions in SilkDB were in that region (BGIBMGA012262 and BGIBMGA012263) but classified as 'antigens'. KAIKObase had

three automated annotations (KAIKOGA048486, KAIKOGA048483 and KAIKOGA048479) but none were accurately reflecting the correct gene model. We annotated the full gene and verified it was a real SNMP homologous, aligning with the *B. mor*i and other Lepidopteran SNMPs. We verified the structure using an EST sequence from KAIKObase (FS809883) spanning the exons of the two automated annotations.

For *H. armiger*a (Lepidoptera) we used Genome data (Pearce et al., 2017) and discovered a new SNMP gene which was orthologous to the *BmorSNMP*3. The automated predictions had three genes spanning that area, but Sanger EST assemblies from InsectaCentral (IC29058AcEcon7899, IC29058AcEcon17308, IC29058AcEcon21585) (Papanicolaou et al., 2005) and RNAseq spanning exon-intron junctions allowed us to annotate the gene model. Further, we verified the gene models of previously reported *SNMP1* [Genbank AAO15604] and *SNMP*2 (Supplementary Table 4).

For *Spodoptera exigua* (Lepidoptera): the *SNMP3* homologue was identified by searching GenBank and reassembling previously reported cDNA sequences (GAFU01000186 and GAFU01009183). Further we identified a partial *SNMP2* gene from GAFU01014218.

For *Plutella xylostella* (Lepidoptera) we found homologues for *SNMP2* and *SNMP3* by manual curation of the previously published *P. xylostella* genome (GCA_000330985.1). The homologue of *SNMP2* was located on scaffold KB207336.1 and was curated with the assistance of transcriptome data (accessions PXUG_V1_014487, PXUG_V1_12791 and RS_pxwb_0023634_c0_s001) available from the genome project databases (http://iae.fafu.edu.cn/DBM) and (http://dbm.dna.affrc.go.jp/px). The homologue of *SNMP3* was located on scaffold KB207313.1 and also supported by RNAseq evidence, which was however partial due to assembly gaps or a misassembly. We used profile, guided alignments with other lepidopteran *SNMP3* genes and the assembled transcriptome (PXUG_V1_028331) to extend the gene to its full length. We further verified the accuracy of this sequence by performing a Trinity de-novo assembly of all the raw RNAseq data: a 2451 bp contig verified the start codon, stop codon, 5' and 3' UTR sequences.

For *Danaus plexippus* (Lepidoptera) we found a total of three full-length SNMPs using genome sequence (GCA_000235995.1). The *SNMP1*, *SNMP2* and *SNMP3* genes were initially annotated from scaffold JH380873, JH384389 and JH390655, respectively. As the scaffold JH384389 was very short (5 kb), we re-annotated using a third version of the *D. plexippus* assembly (scaffolds DPSCF300414, DPSCF300039 and DPSCF300068 respectively) which was available only from MonarchBase directly (http://monarchbase.umassmed.edu). Final full-length sequences were curated, encoding 519 AA, 524 AA and 522 AA, respectively.

For *Heliconius melpomene* (Lepidoptera), we identified and annotated all three SNMPs using the published genome sequence (GCA_000313835.2). The SNMP1 gene was located on scaffold HE671626 and an automated prediction had partially annotated the gene. Using the *Danaus plexippus SNMP1* gene we conducted tBLASTx searches to identify one N-terminus exon but due to gaps, putative misassemblies and lack of RNAseq from relevant tissues, we failed to identify the full C-terminus for this species. The full SNMP2 and SNMP3 genes were similarly curated on scaffolds HE671451 and HE672039 respectively. In the case of *SNMP3*, the gene model was supported by RNAseq evidence.

3.2 Phylogenetic characterization

In the phylogenetic analysis, the insect SNMPs form an independent group from CD36 genes (Vogt et al., 2009). Furthermore, insect SNMPs shows that three distinct clusters are formed: the SNMP1 clade, SNMP2 clade (comprising of the Hymenoptera and Lepidoptera SNMP2) and SNMP3 clade (containing the *Drosophila* SNMP2) (Figure 1). Within each of these clades, each phylum clusters separately, indicative of an evolutionary process that is broadly stable but there are notable exceptions.

An obvious feature in this phylogeny is that what has been called a Lepidopteran SNMP2 is not actually orthologous to the *Drosophila* SNMP2 (Figure 1). When SNMP2 was discovered, researchers named it using the best BLAST score against *Drosophila* SNMP2 rather than undertaking the extensive phylogenetic approach we have done or a functional study as we present herein. Rather the new gene we discover here, which

we named SNMP3, is more likely to be the true orthologue of the *Drosophila* SNMP2. This SNMP3 exists in all Lepidoptera species we surveyed. Within Lepidoptera, we note that for both the SNMP2 and SNMP3, the moth genes form distinct clades and their butterfly orthologues forming a monophyletic outgroup.

3.3 Comparison of time- and spacial-specific expression of SNMP orthologues in insects

To investigate whether SNMP orthologs share similar functional characteristics during development, semi-qRT-PCR was primarily performed to analyze the expression of SNMPs in multiple tissues of *B. mori*, *H. armigera* and *D. melanogaster*, typical insect models of Lepidoptera and Diptera.

By comparing the expression of SNMPs from tissues of both *B. mori* larvae and adults, we found that *BmorSNMP1* and *BmorSNMP3* have more restricted time- and spatial-expression pattern (Figure 2A). As shown in Figure 2A, *BmorSNMP1* was specifically expressed in adult moth antennae. Meanwhile, the expression of *BmorSNMP3* was predominantly detected in larval and adult midgut versus lower expression in male head and slight expression in female head. Conversely, there is a wild range of expression of *BmorSNMP2* in all tested tissues during development.

In comparison of the expression in *B. mori*, similar patterns across tissues and development were found in *H. armigera* SNMPs (Figure 2A). Strikingly, *HarmSNMP3* gene was specifically expressed in both larval and adult midgut. Using RNAseq data from the *H. armigera* Genome Consortium, we found that *HarmSNMP1* was also expressed in male adult tarsus and 5th instar antennae but our semi-qRT-PCR did not capture that. *HarmSNMP2* was also expressed in abdomen, thorax, tarsus, testes, prepupae, and larval foregut, antennae, mouthpart, cuticle and salivary gland besides adult antennae and head (with antennae) according to RNAseq data which is supported from our own PCR result showing broad expression. For *HarmSNMP3*, the RNAseq data showed that it was expressed in the midgut, as supported by our RT-PCR.

Compared with *B. mori* and *H. armigera*, the expression profiles of *SNMP1* ortholog in *Drosophila* is distinct. *DmelSNMP1* and *DmelSNMP2* shared the similar expression characteristic of high expression in multiple tissues of larvae and adults (Figure 2B). Those are supported by previous work on DmelSNMP1 (c.f. FlyBase FBgn0260004 and relevant modENCODE links) that show that it is highly expressed in adults including carcass, eye, fat body, head, hindgut, midgut, salivary gland and larval trachea. *DmelSNMP2* (c.f. FBgn0035815 and relevant modENCODE data) is widely expressed across development but with higher expression in crop and hindgut than in adult carcass, head, larvae hindgut and salivary gland.

We further performed qRT-PCR on BmorSNMPs in larval antennae, larval mouthparts, larval mid-guts, female adult antennae, male adult antennae, adult wings and adult legs. The results are consisted with the semi-qRT-PCR results (Figure 2A). The expression levels of BmorSNMP1 in both female and male antennae are significantly higher than those in other tissues (F=32.64, *P*=0.000) (Figure 2C). The expression levels of BmorSNMP2 are varied in multiple tissues (F=3066, *P*=0.000) (Figure 2D). And BmSNMP3 is remarkably highly expressed in larval mid-gut (F=983.1, *P*=0.000) (Figure 2E).

3.4 SNMPs play a role in immunoresponses

Insect midgut is a major part of the digestive tract and it plays critical roles in immune response, metabolism, homeostasis of electrolytes, osmotic pressure and circulation. CD36 has been shown to have a role in the innate immune response in mouse (Thylur et al., 2017) and SNMP3 belong to CD36 family, so we investigated if BmorSNMP3 was involved in the immune response in the larval midgut. Public microarray data (http://www.silkdb.org/silkdb/) were used to check the expression ratio of genes after the larvae were infected by *Bacillus bombyseptieus*, *Beauveria bassiana*, *Escherichia coli* or *B. mori* nucleopolyhedrovirus (BmNPV). When compared to non-induced control, the ratio of expression for *BmorSNMP2* varied only slightly between types of inoculations and time after inoculation (Figure 2F), an indication that SNMP2 is not involved in immune response. For *BmorSNMP*3, however, the ratio of expression

was notably increased during the first three hours after *B. bombyseptieu*s inoculation and then decreased (Figure 2G). After inoculation with *B. bassiana*, the ratio peaked the 12th hour. After inoculation by BmNPV, the expression ratio was elevated 12 hours post induction and reached its maximum at 24h, which was the total length of the experiment. For *E. coli* there was no variation (Figure 2G). These expression rate changes as a response to infection and that difference across types of infections indicates that SNMP3 is involved to immune responses. Due to lack of any available probe, we could not analyze the expression ratio changes of BmorSNMP1, but we believe it is unlikely that a gene that is specifically expressed in adult antennae is involved in immune response.

3.5 Functional characterization of the silkworm SNMP1 involved in sex pheromone perception

To compare the function of BmorSNMP1 in pheromone detection, pheromone receptor BmorOR1 was also investigated in this study. BmorSNMP1 and BmorOR1 showed similar expression profile in male antennae from 4th day of pupa using RT-PCR (Figure 3A). It provides cue for the time of dsRNA injection.

Behavioural assays were performed within 48 hours after eclosion. However, the rate for non-eclosion pupae is 9-10 % after dsRNA (BmorSNMP1, BmorOR1 or both) injection comparing with 5 % for ddH_2O injection and 0 for control. In one male-female test group, the position of moths was shown (Figure 3B) with the most amount of possible space between them in the dish. Generally, behavioural response could be elicited by female in the majority of male moths. This behaviour ranges from wing vibration and anemotactic walking to female tracking the source of pheromone (Supplementary Video 1). The time for male from recognizing to finding females is postponed after RNAi knocked down of the expression of either BmorSNMP1 or BmorOR1 or both (Figure 3C-E, Supplementary Video 2, Supplementary Table 5). In control, male moths fluttered wings “madly” and showed anemotactic towards the female (Supplementary Video 1). All control (100%) males can recognize and reach the female within 10 min with average time 160 sec (Supplementary Table 5). 91.6±2.8% of

ddH$_2$O treated males could find female in the first 10 min with 250 sec of average time. On the contrary, RNAi treatment results in ≤ 35 % (34.1±5.0% of BmorSNMP1 dsRNA, 25.6±4.6% of BmorOR1 dsRNA and 27.5±4.7% of both) treated males finding female in the first 10 min (F=94.33, *P*=0.000) (Figure 3F, Supplementary Table 5, Supplementary Video 3). The average times of those could locate female after RNAi treatments are around 524 sec (BmorSNMP1$^-$), 620 sec (BmorOR1$^-$) and 641 sec (BmorSNMP1$^-$/BmorOR1$^-$) respectively (Supplementary Table 5). For further comparison of RNAi treated and control males, competing behavioral assays were performed by placing a pair of male moths in the same arena with one female at the longest diameter (Figure 3G and H). Both males exhibit fluttering wings which is a typical sexual behavioural indicating both males detected the pheromone elicited by the female (Supplementary Video 4). However, only control males successfully moved towards and located females within three minutes though the control males did spend almost one minute pursuing the RNAi male. A similar result was found for BmorOR1 RNAi knocked down males with only control moths locating the female within the first three mins (Supplementary Video 5).

After behavioral testing, the expression level of BmorSNMP1 and BmorOR1 in the antennae was detected by real-time PCR ($F_{BmorOR1}$ =50.91, $P_{BmorOR1}$ =0.000; F_{SNMP1} =59.02, P_{SNMP1} =0.000) (Figure 3I). Further BmorSNMP1 dsRNA did not affect the expression of BmorOR1 and vice versa. However, a significant down-expression of BmorSNMP1 and BmorOR1 after RNAi treatment have been found comparing with control and ddH$_2$O injection.

3.6 SNMP1 interacted with olfactory receptors

Previous studies mentioned DmelSNMPs would be involved in olfaction and DmelSNMP1 has a potential interaction with DmelOR22a (German et al., 2013). We hypothesized that there is a crosstalk between insect SNMP1s and pheromone receptors (PR) or Orco homologs in promoting pheromone response. To investigate this, split-ubiquitin yeast hybridization was performed which is considered as a promising method to reveal the interaction between transmembrane proteins. pPR3N vector was

chosen for SNMP1s and BmorOR2 as the prey constructs versus pBT3STE vector for pheromone receptors as bait constructs. The interaction between pNubG-Fe65 and pTSU2-APP was used as positive control (Figure 4A). To exclude the internal interaction between vectors, the interaction of empty constructs of pPR3N and pBT3STE was shown and treated as negative control.

Due to the clear interaction of DmelSNMP1 (CG7000) with DmelOR22a (NM_078729) has been argued, this pair was tested primarily which would provide evidence to whether the yeast system is suitable to our investigation. After 4-day incubation of yeast co-transformed pPR3N-DmelSNMP1 with pBT3STE-DmelOR22a, the positive results have been shown by both SD-Trp-Leu (SD-T/L) and SD-Trp-Leu-His-Ade (SD-T/L/H/A) defective selection (Figure 4B). These results confirmed that DmelSNMP1 interacts with DmelOR22a which has been reported (German et al., 2013), indicating the split-ubiquitin yeast hybridization system could approach our purpose. Subsequently, to evaluate the interaction between SNMP1 orthologs and pheromone receptors in each insect species, the binary pairs of DmelSNMP1/DmelOR67d and BmorSNMP1/BmorOR1 have been detected in yeast hybridization system. The results were suggestive of the interaction of BmorSNMP1/BmorOR1 (Figure 4D) and DmelSNMP1/DmelOR67d (Figure 4F). Moreover, to elucidate whether BmorOR1 and BmorSNMP1 could separately interact with Orco homolog, we further identified co-transformed BmorOR2 (Orco) with BmorOR1 or BmorSNMP1 into yeast. Expectedly, positive results have been revealed (Figure 4C and E). Finally, β-galactosidase activity was evaluated for examining the intensity of the pairwise interaction (Supplementary Table 6). Statistical analysis indicated that there was slight higher but no significant difference in each pair compared with negative control. Taken together, those results inferred that SNMP1 orthologs interact with pheromone receptor/Orco. However, the pairwise interactions are rather weak which did initiate the expressions of HIS3 and ADE2 report genes but couldn't trigger β-galactosidase activity in yeast.

To exclude the unfavorable impact of vectors on fusion protein expression, identical experiments as described as above have been practiced using pPR3N vector for pheromone receptors as the prey constructs versus pBT3STE vector for SNMP1s and BmorOR2 as bait constructs. Consequently, similar results could be found (data not

503 shown). Finally, we proposed a new mode of moth Orco, SNMP1 and pheromone
504 receptor function mechanism (Figure 5). Moth Orco, SNMP1 and pheromone receptor
505 form a heteromer to function when the pheromone compound is delivered by
506 pheromone binding protein in the sensilla (Figure 5).

507

508

4. Discussion

Even though the SNMP subfamily has been known for years, it has not been well studied in Lepidoptera species. The first SNMP was characterized from the wild silkmoth *A. polyphemus* (Rogers et al., 1997). Subsequently, draft whole genome sequencing (WGS) and shallow cDNA sequencing in *D. melanogaster* and other insects allowed the *in-silico* identification of a second protein, named SNMP2, from more Lepidoptera, Diptera, Coleoptera and Hymenoptera (Forstner et al., 2008; Rogers et al., 2001a; Rogers et al., 2001b; Vogt et al., 2009).

Due to the lack of functional work, the naming followed a practice standard in most whole genome sequencing (WGS) projects: first the genes are identified using BLAST searches and, assuming full ascertainment, a draft phylogenetic tree is constructed. Therefore, these phylogenies are often derived on an average linkage clustering (based on sequence identity) or using an unrealistic evolutionary model. Rarely are the assumptions of the phylogeny reconstruction considered. Further, authors rely on bootstrap values as the sole index of accuracy and are rarely interested in investigating deeper. In the case of SNMPs, this network topology guided the assignment of ortholog with gene naming following the convention of serially naming individual sequences as they were described in BLAST searches. Sadly this is still a standard practice in whole genome annotation but new automated procedures such as OrthoDB (Waterhouse et al., 2013) have been made available to facilitate curators. These procedures still assume full ascertainment and an accurate network topology are therefore best suited to be used in conjunction with curation. Further, these approaches neglect the use of expression profiling as a robust approach to assign function.

To date, SNMP1 expression was shown to be highly expressed in pheromone receptor neurons of trichoid sensilla (Forstner et al., 2008; Rogers et al., 2001b). And SNMP1 may directly interact with odorant receptors rather than Orco in those neurons (Benton et al., 2007; German et al., 2013) suggesting SNMP1 may play a specific role in pheromone detection. Subsequent functional studies were limited to *Drosophila* and they showed that SNMP1 is essential for the pheromone receptor (OR67d) to detect the sex pheromone (11 *cis*-vaccenyl acetate) in a native environment (Benton et al., 2007;

539 Gomez-Diaz et al., 2013; Jin et al., 2008). A model was previously proposed from
540 *Drosophila* in which SNMP1 deliver pheromone ligands in the extracellular lymph via an
541 ectodomain tunnel to the cognate pheromone detecting OR in the cilia membrane
542 (Gomez-Diaz et al., 2016). The second protein is yet to be thoroughly characterized that
543 moth SNMP2 is expressed in support cells around odorant sensitive neurons (OSNs) in
544 many olfactory sensilla of antennae (Forstner et al., 2008; Rogers et al., 2001b). Recent
545 study also showed that *Sesamia inferens* SNMP2 is broadly and highly expressed in
546 antennae, legs and wings (Zhang et al., 2013).

547 In this work, we undertook to fully ascertain the SNMP family in the lepidopteran
548 model species the silkworm (*B.mori*) and all other insect with genomic or transcriptomic
549 data. Using all available public data from insects we verified that a novel SNMP clade
550 arose in Lepidoptera. Second, by integrating the expression data with the classic
551 phylogenetic clustering, we show that the fruitfly SNMP2 clusters with the Lepidopteran
552 SNMP3 and *Tribolium* SNMP2, not the Hymenopteran SNMP2. Rather Hymenopteran
553 SNMP2 and Lepidopteran SNMP2 form a distinct monophyletic clade of a gene that is
554 constitutively expressed. Third, based on current expression analysis, moth SNMP1 is
555 restrictively expressed in adult antennae. SNMP2 is broadly expressed in multiple
556 tissues. SNMP3 generally concentrates in larvae and adult midgut indicating the
557 possible function in fatty acid transportation. Further, *B. mori* SNMP3 is also found in
558 adult head (without antennae) and is differentially expressed in whole body when
559 challenged with bacterial infections. On the contrary, the fruitfly SNMPs have similar
560 expression pattern in all tested tissues. Finally, we showed that the silkworm SNMP1 is
561 indeed a functional orthologue of the fruitfly equivalent using RNAi and bioassays. The
562 BmorSNMP1 RNAi treated male moths took much longer time to find the female, which
563 is consistent with a study in *Drosophila* SNMP1. It was reported that *snmp1* mutated
564 *Drosophila* did not abolish cVA responsiveness or cause high spontaneous activity. The
565 cVA responses in *snmp1* mutated fly exhibited a delayed onset and took longer to reach
566 peak activity than wild-type fly. Furthermore, loss of SNMP1 caused a dramatic delay in
567 signal termination (Li et al., 2014).. After the pheromone receptor, BmorOR1, was
568 introduced in OR67d ORNs of *snmp1* mutated flies, it was found that the ORNs showed
569 slow activation and deactivation kinetics in response to bombykol (Li et al., 2014). When

BmorOR1 was expressed in *Xenopus* oocytes, the presence of BmorSNMP1 accelerated receptor activation and deactivation, suggesting BmorSNMP1 plays a critical role in rapid kinetics of the pheromone response in insects (Li et al., 2014). For the first time, we showed SNMP1 does not only bind to pheromone receptor (BmorOR1), but also the co-receptor, BmorOrco. Based on this discovery, we brought out a new mode of insect pheromone receptor functional complex in silkworm, which is composed of BmorOrco, BmorSNMP1 and BmorOR1. Our results provide new evidence for SNMPs function *in vivo*. These can be readily applied to other insects that can mate in a laboratory setting. This study also built a model by using phylogenomics plus functional characterization approaches to study proteins in vitro and in vivo.

Acknowledgments

We thank Dr. Richard Vogt for comments and discussion, Dr. Richard D. Newcomb and Dr. Selene van der Poel for poviding constructs of DmelOR22a, DmelOR67d and DmelSNMP1; Dr. Ting-Cai Cheng for assisting microarray data analysis, Dr. Faisal Younus and Dr. Olivia Leitch for providing *Drosophila* strain and assisting with dissections, and Dr. Chengzhi Chen for supporting statistical analysis. This work was supported by China Postdoctoral Science Foundation (2013M531930), partially funded by National Program on Key Basic Research Project (973 Program, 2012CB114600) and funding from the Transformational Biology Capability Platform (TCP1: Dissecting Adaptive Potential) of the Commonwealth Scientific and Industrial Research Organisation, Australia. Dr Wei Xu is the recipient of an Australian Research Council Discovery Early Career Researcher Award (DECRA) (DE160100382).

Author Contributions

H.-J.Z. conceived, designed and investigated the experiments, analyzed the data and wrote the manuscript. W.X. performed some data analysis, revised and edited the manuscript. Q.-M.C. and L.-N.S. detected the expression of BmorSNMPs and part of behavior experiments on *B. mori*. Q.-Y.X. supervised and co-designed the work and revised the manuscript. A.P. co-ordinated the genomic research, co-designed the work, supervised the analysis, data analysis, drafted and revised the manuscript. A. A. co-supervised the insect bioassay research, funded the lab for *Drosophila* and *H. armigera* experiments. All authors contributed to the manuscript, read and approved the final version. All authors report no conflicts of interest.

Figure Legends

Figure 1. Phylogenetic analysis of insect SNMP homologs. ML tree of insect SNMPs built by IQ-TREE. Blue, green and purple branches present SNMP1, SNMP2 and SNMP3. The bootstrap is 1000 replications. To simplify, abbreviation name for each insect species are listed as below: *H. armigera*, Harm; *H. assulta*, Hass; *Heliothis virescens*, Hvir; *Mamestra brassiae,* Mbra; *Antheraea polyphemus*, Apol; *B. mori*, Bmor; *Samia cynthia ricini*, Scyn; *Manduca sexta*, Msex; *Chilo suppressalis*, Csup; *Cnaphalocrocis medinalis*, Cmed; *Ostrinia nubilalis*, Onub; *Ostrinia furnacalis*, Ofur; *Plutella xylostella*, Pxyl; *Agrotis ipsilon*, Aips; *Danaus plexippus*, Dple; *Spodoptera litura*, Slit; *S. frugiperda*, Sfru; *Sesamia inferens*, Sinf; *Chrysopa pallens*, Cpal; *D. melanogaster*, Dmel; *D. pseudoobscura*, Dpse; *Anopheles gambiae*, Agam; *Dendroctonus ponderosae*, Dpon; *Apis mellifera*, Amel; *Apis florea,* Aflo; *A. pisum*, Apis; *Solenopsis Invicta*, Sinv; *Megachile rotundata*, Mrot; *Bombus terrestris*, Bter; *Cimex lectularius*, Clec; *Nasonia vitripennis*, Nvit; *G. morsitans*, Gmor; *Ips typographus*, Ityp; *P. humanus*, Phum; *Culex quinquefasciatus*, Cqui, and *T. castaneum*, Tcas,

Figure 2. Specific tissue and time expression pattern of SNMPs. A, Gene expression of *B. mori* and *H. armigera* SNMP orthologs in multiple tissues by using RT-PCR. **B**, Expression of DmelSNMPs in *Drosophila* larval and adult tissues. **C**, **D** and **E**, the relative expression of BmorSNMP1, 2 and 3 in various larvae and adult tissues by real time PCR (n=3). ANOVA analysis followed by Tukey's multiple comparisons test was used for pairwise comparison. The significant differences have been indicated as follows, otherwise no significant differences. a, antennae (F) vs. each other group, P=0.000; b, antennae (M) vs. each other group, P=0.001; c, antennae (F) vs. antennae (M), P=0.024; d, antennae (L) vs. each other group, P=0.000; f, mouthparts (L) vs. wings (A), P=0.001; g, mouthparts (L) vs. legs (A), P=0.002; h, mid-gut (L) vs. wings (A), P=0.037; i, mid-gut (L) vs. legs (A), P=0.017; mid-guts (L) vs. each other group, P=0.000. **F** and **G**, *in silico* expression ratio analysis of BmorSNMP2 and 3 in individual larva infested by bacteria or virus against control treated by same volume of ddH_2O.

641 The larvae at 3rd day of 5th instar was infected by *Bacillus bombyseptieus* (red line with
642 hollow circles), *Beauveria bassiana* (blue line with hollow circles), *Escherichia coli*
643 (green line with hollow triangles) or *B. mori* nucleopolyhedrovirus (BmNPV) (black line
644 with hollow triangles) for three hours and then fed by normal artificial food. The whole
645 insects were collected at different time for oligonucleotide arrays. The expression ratio
646 of BmorSNMP2 and 3 were screened through microarray data using gene ID in genome.

647

648 **Figure 3. Behaviour assays on SNMP1/OR1 knock-down moth treated by RNAi. A**,
649 Expression of BmorSNMP1 and BmorOR1 in male antennae during the last four days of
650 pupae and the 1st day of the eclosion. **B**, Schematic diagram of one male-female moth
651 recognizing and finding assays. **C**, Positions for single pair of male-female moth before
652 and after 20 min. **D**, The ratios of male moths approached female moths at various time
653 points. Statistic analyzed the behavioural data for each male approaching female which
654 indicated by survival curves. Log-rank analysis was used and showed the significant
655 differences between curves (χ^2 =244.9, p<0.001). **E**, the times male moths spent to
656 find females (n=100). ANOVA analysis also showed the significant differences between
657 groups (F=118.4, p<0.001). Then Tukey's multiple comparisons test was used for
658 pairwise comparison. I: control vs. ddH_2O; II: control vs. BmorOR1$^+$/BmorSNMP1$^-$; III:
659 control vs. BmorOR1$^-$/BmorSNMP1$^+$; IV: control vs. BmorOR1$^-$/BmorSNMP1$^-$; V: ddH_2O
660 vs. BmorOR1$^+$/BmorSNMP1$^-$; VI: ddH_2O vs. BmorOR1$^-$/BmorSNMP1$^+$; VII: ddH_2O vs.
661 BmorOR1$^-$/BmorSNMP1$^-$; VIII: BmorOR1$^+$/BmorSNMP1$^-$ vs. BmorOR1$^-$/BmorSNMP1$^+$;
662 IX: BmorOR1$^+$/BmorSNMP1$^-$ vs. BmorOR1$^-$/BmorSNMP1$^-$; X: BmorOR1$^-$/BmorSNMP1$^+$
663 vs. BmorOR1$^-$/BmorSNMP1$^-$. a, P_I=0.677; b-g, P_{II-VII}=0.000;h, P_{VIII}=0.432; i, P_{IX}=0.753; j,
664 P_X=0.901. **F**, The mating ratios (%) of RNAi treated male moths in the first 10 mins.
665 Tukey's multiple comparisons test was used for pairwise comparison (n=100). a,
666 P_I=0.504; b-g, P_{II-VII}=0.000;h, P_{VIII}=0.530; i, P_{IX}=0.749; j, P_X=0.999. **G** and **H**, Photo
667 and trace examples for competing behavioural assay. Red and blue round bars present
668 intact (control) and RNAi treated male moth respectively. Red and blue dot lines
669 indicate moving paths of moth getting to female in 3 min. Hollow and hard arrows
670 respectively suggest start-stop positions. **I**, Relative expression level of BmorSNMP1

and BmorOR1 in male moths' antennae after RNAi were detected by real-time PCR (n=3). Pairwise comparisons of BmorOR1 expression are as follows. a, P_{I}=0.007; b, P_{II}=0.317; c-d, f-l, $P_{III\text{-}IV;\ VI\text{-}IX}$=0.000; e, P_{V}=0.866; j, P_{X}=0.999. Pairwise comparisons of BmorSNMP1 expression are as follows. k, P_{I}=0.071; l, n, o, q, r, t, $P_{II,\ V,\ VII,\ VIII,\ X,}$ =0.000; m, P_{III}=0.867; p, P_{VI}=0.703; s, P_{IX}=0.980. Error bar showed the standard error of the mean (SEM).

Figure 4. The interaction between SNMP1 and pheromone receptor from *D. melanogaster* and *B. mori* using split-ubiquitin yeast hybridization system. After co-transformed two fusion protein constructs into yeast, positive clones could be shown in amino acid defective selection medium (SD-T/L or SD-T/L/H/A) if the two target proteins interacted with each other. **A**, two pairs of pNubG-Fe65 x pTSU2-APP and pPR3N x pBT3STE were used as positive and negative controls respectively. **B**, the reported interaction of DmelSNMP1 and DmelOR22a was examined by yeast hybridization assays. **C**, the interactions between the BmorOR2 (Orco) and pheromone receptor (BmorOR1). **D** and **E**, the potential interactions of SNMP1 orthologs with pheromone receptors and Orco in *B. mori* were separately evaluated by identical treatment. F, the interactions between the DmelSNMP1 and DmelOR67d.

Figure 5. The new mode of Orco, SNMP1 and pheromone receptor function mechanism. *In B. mori*, pheromone sensilla, sex pheromone compound (bombykol) was transported by pheromone binding protein (PBP) to the functional complex which is composed by pheromone receptor (BmorOR1), co-receptor (BmorOrco) and BmorSNMP1, to activate the moth responses.

695 **Table 1**. **Identification of insect SNMPs**

Gene name	New name	Species	Insect oder	High-confidence	Edited or new	Genbank
SexiSNMP2		*Spodoptera exigua*	Lepidoptera		Partial public	GAFU01014218
CpalSNMP1		*Chrysopa pallens*	Neuroptera	Yes	Partial public	GAGF01037985
CpalSNMP2		*Chrysopa pallens*	Neuroptera	Yes	Partial public	GAGF01038012
CsupSNMP3		*Chilo suppressalis*	Lepidoptera	yes	Partial public	GAJS01010313
AmelSNMP2		*Apis mellifera*	Hymenoptera	Yes	Edited	GB12830-PA
DmelSNMP2		*Drosophila melanogaster*	Diptera	Yes		NP_001036593
DpseSNMP2		*Drosophila pseudoobscura*	Diptera	Yes	Edited	XP_002024979
PhumSNMP1		*Pediculus humanus corporis*	Psocodea	Yes	New	XP_002432887.1
PhumSNMP2		*Pediculus humanus corporis*	Psocodea	Yes	New	
GmorSNMP1		*Glossina morsitans*	Diptera	Yes	New	
GmorSNMP2		*Glossina morsitans*	Diptera	Yes	New	
ApisSNMP1		*Acyrthosiphon pisum*	Hemiptera	Yes	New	
ClecSNMP1		*Cimex lectularius*	Hemiptera	Yes	New	
ClecSNMP2a		*Cimex lectularius*	Hemiptera	Yes		XP_014253859.1
ClecSNMP2b		*Cimex lectularius*	Hemiptera	Yes		XP_014239380.1
SinvSNMP1a		*Solenopsis invicta*	Hymenoptera	Yes		XP_011159295.1
SinvSNMP1b		*Solenopsis invicta*	Hymenoptera	Yes		XP_011159294.1
BmorSNMP3		*Bombyx mori*	Lepidoptera	Yes		XP_004933211.1
DpleSNMP1		*Danaus plexippus*	Lepidoptera	Yes		OWR54175.1
DpleSNMP2		*Danaus plexippus*	Lepidoptera	Yes	New	
DpleSNMP3		*Danaus plexippus*	Lepidoptera	Yes	New	
HarmSNMP2		*Helicoverpa armigera*	Lepidoptera	Yes	New	
HarmSNMP3		*Helicoverpa armigera*	Lepidoptera	Yes	New	XP_021184873.1
HmelSNMP1		*Heliconius melpomene*	Lepidoptera	Yes	New	
HmelSNMP2		*Heliconius melpomene*	Lepidoptera	Yes	New	
HmelSNMP3		*Heliconius melpomene*	Lepidoptera	Yes	New	
PxylSNMP2		*Plutella xylostella*	Lepidoptera	Yes	New	

PxylSNMP3	*Plutella xylostella*	Lepidoptera	Yes	New	
SexiSNMP3	*Spodoptera exigua*	Lepidoptera		New	AKT26505.1
MsexSNMP2	*Manduca sexta*	Lepidoptera			AAG49365
MsexSNMP1	*Manduca sexta*	Lepidoptera			AAG49366
HarmSNMP1	*Helicoverpa armigera*	Lepidoptera	Yes		AAO15604
OfurSNMP2	*Ostrinia furnacalis*	Lepidoptera			ADQ73891
OfurSNMP1	*Ostrinia furnacalis*	Lepidoptera			ADQ73894
CquiSNMP1a	*Culex quinquefasciatus*	Diptera	Yes		AEK32386
CquiSNMP1b	*Culex quinquefasciatus*	Diptera	Yes		AEK32387
CquiSNMP1c	*Culex quinquefasciatus*	Diptera	Yes		AEK32388
CquiSNMP2	*Culex quinquefasciatus*	Diptera	Yes		AEK32389
CmelSNMP1	*Cnaphalocrocis medinalis*	Lepidoptera			AFG73002
CmelSNMP2	*Cnaphalocrocis medinalis*	Lepidoptera			AFG73003
CsupSNMP1	*Chilo suppressalis*	Lepidoptera			AFS50073
CsupSNMP2	*Chilo suppressalis*	Lepidoptera			AFS50074
SlitSNMP1	*Spodoptera littoralis*	Lepidoptera			
SlitSNMP2	*Spodoptera littoralis*	Lepidoptera			
SlittSNMP1	*Spodoptera litura*	Lepidoptera			AGN48098
SlittSNMP2	*Spodoptera litura*	Lepidoptera			AGN48099
BmorSNMP2	*Bombyx mori*	Lepidoptera	Yes		AK383140
ApolSNMP2	*Antheraea polyphemus*	Lepidoptera			AM905329
HvirSNMP2	*Heliothis virescens*	Lepidoptera			B2RFN2
BmorSNMP1	*Bombyx mori*	Lepidoptera	Yes		CAB65730
OnubSNMP1	*Ostrinia nubilalis*	Lepidoptera			E5EZW6
OnubSNMP2	*Ostrinia nubilalis*	Lepidoptera			E5EZW9
ItypSNMP1	*Ips typographus*	Coleoptera			GACR01000057
ItypSNMP2	*Ips typographus*	Coleoptera			GACR01000058
PxylSNMP1	*Plutella xylostella*	Lepidoptera	Yes		HM536984
AipsSNMP1	*Agrotis ipsilon*	Lepidoptera			KC182740
AipsSNMP2	*Agrotis ipsilon*	Lepidoptera			KC182741

SinfSNMP1		*Sesamia inferens*	Lepidoptera		KC907737
SinfSNMP2		*Sesamia inferens*	Lepidoptera		KC907738
DmelSNMP1		*Drosophila melanogaster*	Diptera	Yes	NP_650953
ApolSNMP1		*Antheraea polyphemus*	Lepidoptera		O02351
HvirSNMP1		*Heliothis virescens*	Lepidoptera		Q9U1G3
DpseSNMP1		*Drosophila pseudoobscura*	Diptera	Yes	XP_001359654
NvitSNMP1-like	NvitSNMP1a	*Nasonia vitripennis*	Hymenoptera	Yes	XP_001606602
NvitSNMP1-like	NvitSNMP1c	*Nasonia vitripennis*	Hymenoptera	Yes	XP_001606675
NvitSNMP1-like	NvitSNMP1b	*Nasonia vitripennis*	Hymenoptera	Yes	XP_001606682
NvitSNMP1-like	NvitSNMP1d	*Nasonia vitripennis*	Hymenoptera	Yes	XP_001606692
TcasSNMP1		*Tribollium castaneum*	Coleoptera	Yes	XP_001816436
BterSNMP1-like	BterSNMP2	*Bombus terrestris*	Hymenoptera		XP_003395247
BterSNMP1-like	BterSNMP1	*Bombus terrestris*	Hymenoptera		XP_003398464
AfloSNMP1		*Apis florea*	Hymenoptera		XP_003691612
AfloSNMP2		*Apis florea*	Hymenoptera		XP_003694668
MrotSNMP1		*Megachile rotundata*	Hymenoptera		XP_003699727
MrotSNMP2		*Megachile rotundata*	Hymenoptera		XP_003703897
AgamSNMP1		*Anopheles gambiae*	Diptera	Yes	XP_312496
AgamSNMP2		*Anopheles gambiae*	Diptera	Yes	XP_315733
AmelSNMP1		*Apis mellifera*	Hymenoptera	Yes	XP_397430
TcasSNMP2		*Tribollium castaneum*	Coleoptera	Yes	XP_970008
DponSNMP1		*Dendroctonus ponderosae*	Coleoptera	Yes	JQ855703
DponSNMP1a		*Dendroctonus ponderosae*	Coleoptera	Yes	KC113423
DponSNMP2		*Dendroctonus ponderosae*	Coleoptera	Yes	KC113436

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402.

Benton, R., Vannice, K.S., Vosshall, L.B., 2007. An essential role for a CD36-related receptor in pheromone detection in *Drosophila*. Nature 450, 289-293.

Cheng, T., Lin, P., Huang, L., Wu, Y., Jin, S., Liu, C., Xia, Q., 2016. Genome-wide analysis of host responses to four different types of microorganisms in *Bombyx mori* (Lepidoptera: Bombycidae). J Insect Sci 16.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792-1797.

Fetchko, M., Stagljar, I., 2004. Application of the split-ubiquitin membrane yeast two-hybrid system to investigate membrane protein interactions. Methods 32, 349-362.

Forstner, M., Gohl, T., Gondesen, I., Raming, K., Breer, H., Krieger, J., 2008. Differential expression of SNMP-1 and SNMP-2 proteins in pheromone-sensitive hairs of moths. Chem. Senses 33, 291-299.

German, P.F., van der Poel, S., Carraher, C., Kralicek, A.V., Newcomb, R.D., 2013. Insights into subunit interactions within the insect olfactory receptor complex using FRET. Insect Biochem. Mol. Biol. 43, 138-145.

Gomez-Diaz, C., Bargeton, B., Abuin, L., Bukar, N., Reina, J.H., Bartoi, T., Graf, M., Ong, H., Ulbrich, M.H., Masson, J.F., Benton, R., 2016. A CD36 ectodomain mediates insect pheromone detection via a putative tunnelling mechanism. Nat Commun 7.

Gomez-Diaz, C., Reina, J.H., Cambillau, C., Benton, R., 2013. Ligands for pheromone-sensing neurons are not conformationally activated odorant binding proteins. PLoS biol. 11, e1001546.

Gu, S.H., Yang, R.N., Guo, M.B., Wang, G.R., Wu, K.M., Guo, Y.Y., Zhou, J.J., Zhang, Y.J., 2013. Molecular identification and differential expression of sensory neuron membrane proteins in the antennae of the black cutworm moth *Agrotis ipsilon*. J Insect Physiol 59, 430-443.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59, 307-321.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35, 518-522.

Jin, X., Ha, T.S., Smith, D.P., 2008. SNMP is a signaling component required for pheromone sensitivity in *Drosophila*. Proc Natl Acad Sci U S A 105, 10996-11001.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14, 587-+.

Kirkness, E.F., Haas, B.J., Sun, W.L., Braig, H.R., Perotti, M.A., Clark, J.M., Lee, S.H., Robertson, H.M., Kennedy, R.C., Elhaik, E., Gerlach, D., Kriventseva, E.V., Elsik, C.G., Graur, D., Hill, C.A., Veenstra, J.A., Walenz, B., Tubio, J.M.C., Ribeiro, J.M.C., Rozas, J., Johnston, J.S., Reese, J.T., Popadic, A., Tojo, M., Raoult, D., Reed, D.L., Tomoyasu, Y., Krause, E., Mittapalli, O., Margam, V.M., Li, H.M., Meyer, J.M., Johnson, R.M., Romero-Severson, J., VanZee, J.P., Alvarez-Ponce, D., Vieira, F.G., Aguade, M., Guirao-Rico, S., Anzola, J.M., Yoon, K.S., Strycharz, J.P., Unger, M.F., Christley, S., Lobo, N.F., Seufferheld, M.J., Wang, N.K., Dasch, G.A., Struchiner, C.J., Madey, G., Hannick, L.I., Bidwell, S., Joardar, V., Caler, E., Shao, R.F., Barker, S.C., Cameron, S., Bruggner, R.V., Regier, A., Johnson, J., Viswanathan, L., Utterback, T.R., Sutton, G.G., Lawson, D., Waterhouse, R.M., Venter, J.C., Strausberg, R.L., Berenbaum, M.R., Collins, F.H., Zdobnov, E.M., Pittendrigh, B.R., 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. P Natl Acad Sci USA 107, 12168-12173.

Li, P.Y., Qin, Y.C., 2011. Molecular cloning and characterization of sensory neuron membrane protein and expression pattern analysis in the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). Appl Entomol Zool 46, 497-504.

Li, Z.Z., Ni, J.F.D., Huang, J., Montell, C., 2014. Requirement for *Drosophila* SNMP1 for Rapid Activation and Termination of Pheromone-Induced Activity. Plos Genet 10.

Liu, C.C., Zhang, J., Liu, Y., Wang, G.R., Dong, S.L., 2014. Expression of snmp1 and snmp2 genes in antennal sensilla of *Spodoptera exigua* (Hubner). Arch Insect Biochem 85, 114-126.

Liu, S., Qiao, F., Liang, Q.M., Huang, Y.J., Zhou, W.W., Gong, Z.J., Cheng, J., Zhu, Z.R., 2013a. Molecular characterization of two sensory neuron membrane proteins from *Chilo suppressalis* (Lepidoptera: Pyralidae). Ann Entomol Soc Am 106, 378-384.

Liu, S., Zhang, Y.R., Zhou, W.W., Liang, Q.M., Yuan, X., Cheng, J.A., Zhu, Z.R., Gong, Z.J., 2013b. Identification and characterization of two sensory neuron membrane proteins from *Cnaphalocrocis medinalis* (Lepidoptera: Pyralidae). Arch Insect Biochem 82, 29-42.

Liu, Y., Gu, S., Zhang, Y., Guo, Y., Wang, G., 2012. Candidate olfaction genes identified within the *Helicoverpa armigera* antennal transcriptome. Plos One 7, e48260.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods 25, 402-408.

Moghani-Ghoroghi, F., Moshkdanian, G., Sehat, M., Nematollahi-Mahani, S.N., Ragerdi-Kashani, I., Pasbakhsh, P., 2018. Melatonin pretreated blastocysts along with calcitonin administration improved implantation by upregulation of heparin

777 binding-epidermal growth factor expression in murine endometrium. Cell J 19 (4),
778 599-606.

779 Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and
780 effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol
781 Biol Evol 32, 268-274.

782 Nichols, Z., Vogt, R.G., 2008. The SNMP/CD36 gene family in Diptera, Hymenoptera
783 and Coleoptera: *Drosophila melanogaster*, *D. pseudoobscura*, *Anopheles*
784 *gambiae*, *Aedes aegypti*, *Apis mellifera*, and *Tribolium castaneum*. Insect
785 Biochem. Mol. Biol. 38, 398-415.

786 Papanicolaou, A., Joron, M., Mcmillan, W.O., Blaxter, M.L., Jiggins, C.D., 2005.
787 Genomic tools and cDNA derived markers for butterflies. Mol Ecol 14, 2883-2897.

788 Pearce, S.L., Clarke, D.F., East, P.D., Elfekih, S., Gordon, K.H.J., Jermiin, L.S.,
789 McGaughran, A., Oakeshott, J.G., Papanikolaou, A., Perera, O.P., Rane, R.V.,
790 Richards, S., Tay, W.T., Walsh, T.K., Anderson, A., Anderson, C.J., Asgari, S.,
791 Board, P.G., Bretschneider, A., Campbell, P.M., Chertemps, T., Christeller, J.T.,
792 Coppin, C.W., Downes, S.J., Duan, G., Farnsworth, C.A., Good, R.T., Han, L.B.,
793 Han, Y.C., Hatje, K., Horne, I., Huang, Y.P., Hughes, D.S.T., Jacquin-Joly, E.,
794 James, W., Jhangiani, S., Kollmar, M., Kuwar, S.S., Li, S., Liu, N.Y., Maibeche,
795 M.T., Miller, J.R., Montagne, N., Perry, T., Qu, J., Song, S.V., Sutton, G.G., Vogel,
796 H., Walenz, B.P., Xu, W., Zhang, H.J., Zou, Z., Batterham, P., Edwards, O.R.,
797 Feyereisen, R., Gibbs, R.A., Heckel, D.G., McGrath, A., Robin, C., Scherer, S.E.,
798 Worley, K.C., Wu, Y.D., 2017. Genomic innovations, transcriptional plasticity and
799 gene loss underlying the evolution and divergence of two highly polyphagous and
800 invasive *Helicoverpa* pest species. BMC biology 15, 63.

801 Rogers, M.E., Krieger, J., Vogt, R.G., 2001a. Antennal SNMPs (sensory neuron
802 membrane proteins) of Lepidoptera define a unique family of invertebrate CD36-
803 like proteins. J. Neurobiol. 49, 47-61.

804 Rogers, M.E., Steinbrecht, R.A., Vogt, R.G., 2001b. Expression of SNMP-1 in olfactory
805 neurons and sensilla of male and female antennae of the silkmoth *Antheraea*
806 *polyphemus*. Cell Tissue Res. 303, 433-446.

807 Rogers, M.E., Sun, M., Lerner, M.R., Vogt, R.G., 1997. Snmp-1, a novel membrane
808 protein of olfactory neurons of the silk moth *Antheraea polyphemus* with homology
809 to the CD36 family of membrane proteins. J. Biol. Chem. 272, 14792-14799.

810 Scheper, W., Thaminy, S., Kais, S., Stagljar, I., Romisch, K., 2003. Coordination of N-
811 glycosylation and protein translocation across the endoplasmic reticulum
812 membrane by Sss1 protein. J. Biol. Chem. 278, 37998-38003.

813 Snider, J., Kittanakom, S., Curak, J., Stagljar, I., 2010. Split-ubiquitin based membrane
814 yeast two-hybrid (MYTH) system: a powerful tool for identifying protein-protein
815 interactions. J Vis Exp.

816 Stagljar, I., Korostensky, C., Johnsson, N., te Heesen, S., 1998. A genetic system
817 based on split-ubiquitin for the analysis of interactions between membrane
818 proteins in vivo. P Natl Acad Sci USA 95, 5187-5192.

Thaminy, S., Auerbach, D., Arnoldo, A., Stagljar, I., 2003. Identification of novel ErbB3-interacting factors using the split-ubiquitin membrane yeast two-hybrid system. Genome Res 13, 1744-1753.

Thylur, R.P., Wu, X.Z., Gowda, N.M., Punnath, K., Neelgund, S.E., Febbraio, M., Gowda, D.C., 2017. CD36 receptor regulates malaria-induced immune responses primarily at early blood stage infection contributing to parasitemia control and resistance to mortality. J. Biol. Chem. 292, 9394-9408.

Vogt, R.G., Miller, N.E., Litvack, R., Fandino, R.A., Sparks, J., Staples, J., Friedman, R., Dickens, J.C., 2009. The insect SNMP gene family. Insect Biochem. Mol. Biol. 39, 448-456.

Waterhouse, R.M., Tegenfeldt, F., Li, J., Zdobnov, E.M., Kriventseva, E.V., 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res 41, D358-D365.

Zhang, J., Liu, Y., Walker, W.B., Dong, S.L., Wang, G.R., 2015. Identification and localization of two sensory neuron membrane proteins from *Spodoptera litura* (Lepidoptera: Noctuidae). Insect Sci 22, 399-408.

Zhang, Y.N., Jin, J.Y., Jin, R., Xia, Y.H., Zhou, J.J., Deng, J.Y., Dong, S.L., 2013. Differential expression patterns in chemosensory and non-chemosensory tissues of putative chemosensory genes identified by transcriptome analysis of insect pest the purple stem borer *Sesamia inferens* (Walker). Plos One 8.

Legends for Supplementary Material

Supplementary Resource 1 Video of single pair of male-female behavioural assays

Supplementary Resource 2 Example for those males lost sexual behavioural response elicited by female after RNAi

Supplementary Resource 3 Video of single pair of male-female behavioural assays. Male moth went away though male and female were placed close to each other directly.

Supplementary Resource 4 Video of competition behavioural assay after BmorSNMP1 dsRNA injection

Supplementary Resource 5 Video of competition behavioural assay after BmorOR1 dsRNA injection

Supplementary Table 1 Primers used in semi-qRT-PCR are listed as below

Supplementary Table 2 Primers used for split-ubiquitin yeast hybridization are listed as below.

Supplementary Table 3 Curated SNMPs within the Arthropoda

Supplementary Table 4 Other known SNMPs used in the phylogenetic analysis

Supplementary Table 5 Percentage of male moths which can recognize and find female in time frames.

Supplementary Table 6 The data of β-GAL experiments for pairwise interaction of proteins.

Figure 1

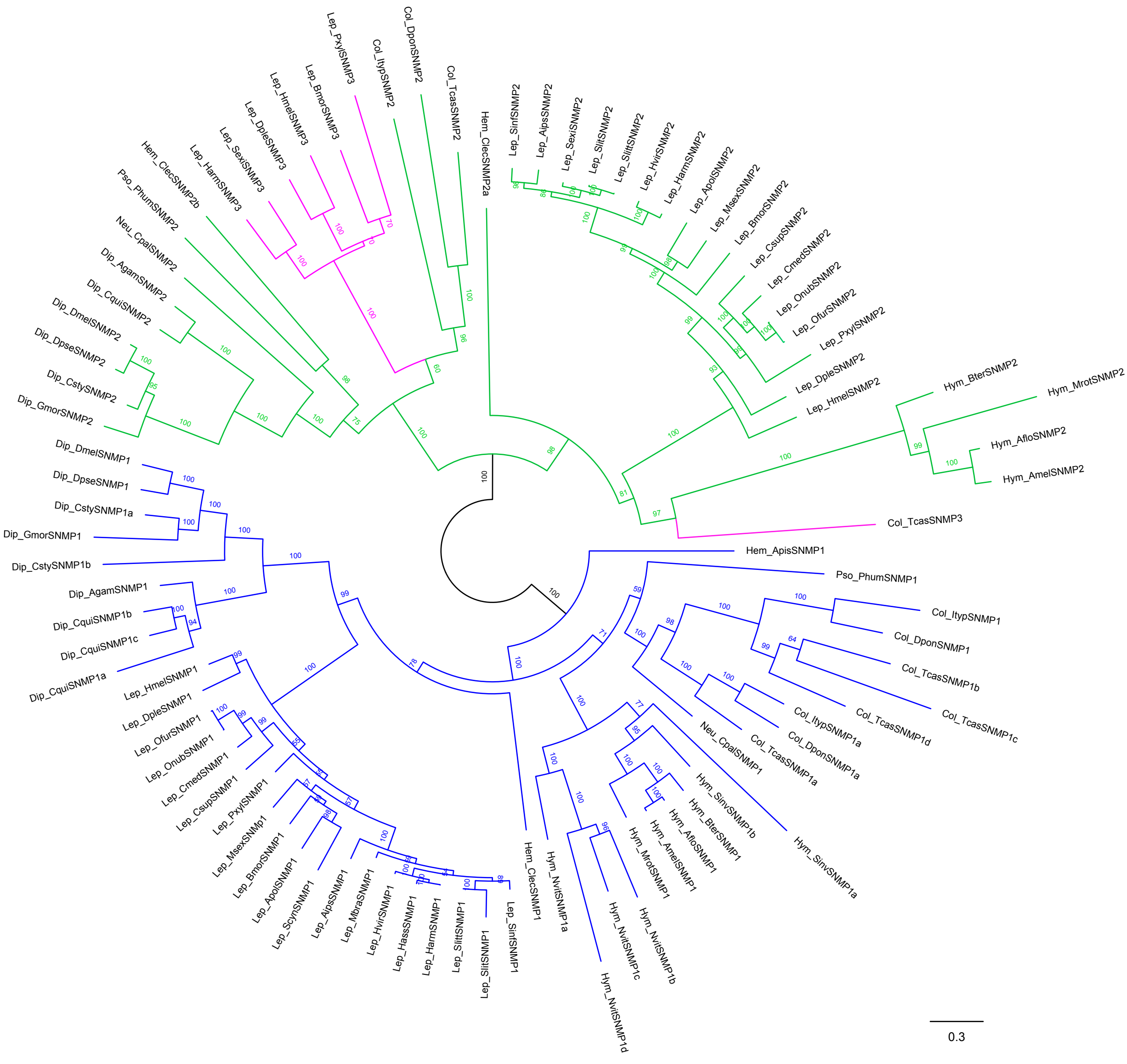

Lep_PxylSNMP3
Col_ItypSNMP2
Col_DponSNMP2
Col_TcasSNMP2
Lep_BmorSNMP3
Lep_HmelSNMP3
Lep_DpleSNMP3
Lep_SexiSNMP3
Lep_HarmSNMP3
Hem_ClecSNMP2b
Pso_PhumSNMP2
Neu_CpalSNMP2
Dip_AgamSNMP2
Dip_CquiSNMP2
Dip_DmelSNMP2
Dip_DpseSNMP2
Dip_CstySNMP2
Dip_GmorSNMP2
Hem_ClecSNMP2a
Lep_SinfSNMP2
Lep_AipsSNMP2
Lep_SexiSNMP2
Lep_SlitSNMP2
Lep_SlittSNMP2
Lep_HvirSNMP2
Lep_HarmSNMP2
Lep_ApolSNMP2
Lep_MsexSNMP2
Lep_BmorSNMP2
Lep_CsupSNMP2
Lep_CmedSNMP2
Lep_OnubSNMP2
Lep_OfurSNMP2
Lep_PxylSNMP2
Lep_DpleSNMP2
Lep_HmelSNMP2
Hym_BterSNMP2
Hym_MrotSNMP2
Hym_AfloSNMP2
Hym_AmelSNMP2
Col_TcasSNMP3
Hem_ApisSNMP1
Pso_PhumSNMP1
Col_ItypSNMP1
Col_DponSNMP1
Col_TcasSNMP1b
Col_TcasSNMP1d
Col_TcasSNMP1c
Col_ItypSNMP1a
Col_DponSNMP1a
Col_TcasSNMP1a
Neu_CpalSNMP1
Hym_SinvSNMP1b
Hym_BterSNMP1
Hym_AfloSNMP1
Hym_AmelSNMP1
Hym_MrotSNMP1
Hym_SinvSNMP1a
Hym_NvitSNMP1c
Hym_NvitSNMP1b
Hym_NvitSNMP1d
Hym_NvitSNMP1a
Hem_ClecSNMP1
Lep_SinfSNMP1
Lep_SlitSNMP1
Lep_HarmSNMP1
Lep_HassSNMP1
Lep_HvirSNMP1
Lep_MbraSNMP1
Lep_AipsSNMP1
Lep_ScynSNMP1
Lep_ApolSNMP1
Lep_BmorSNMP1
Lep_MsexSNMp1
Lep_PxylSNMP1
Lep_CsupSNMP1
Lep_CmedSNMP1
Lep_OnubSNMP1
Lep_OfurSNMP1
Lep_DpleSNMP1
Lep_HmelSNMP1
Dip_CquiSNMP1a
Dip_CquiSNMP1c
Dip_CquiSNMP1b
Dip_AgamSNMP1
Dip_CstySNMP1b
Dip_GmorSNMP1
Dip_CstySNMP1a
Dip_DpseSNMP1
Dip_DmelSNMP1
0.3

Figure 2

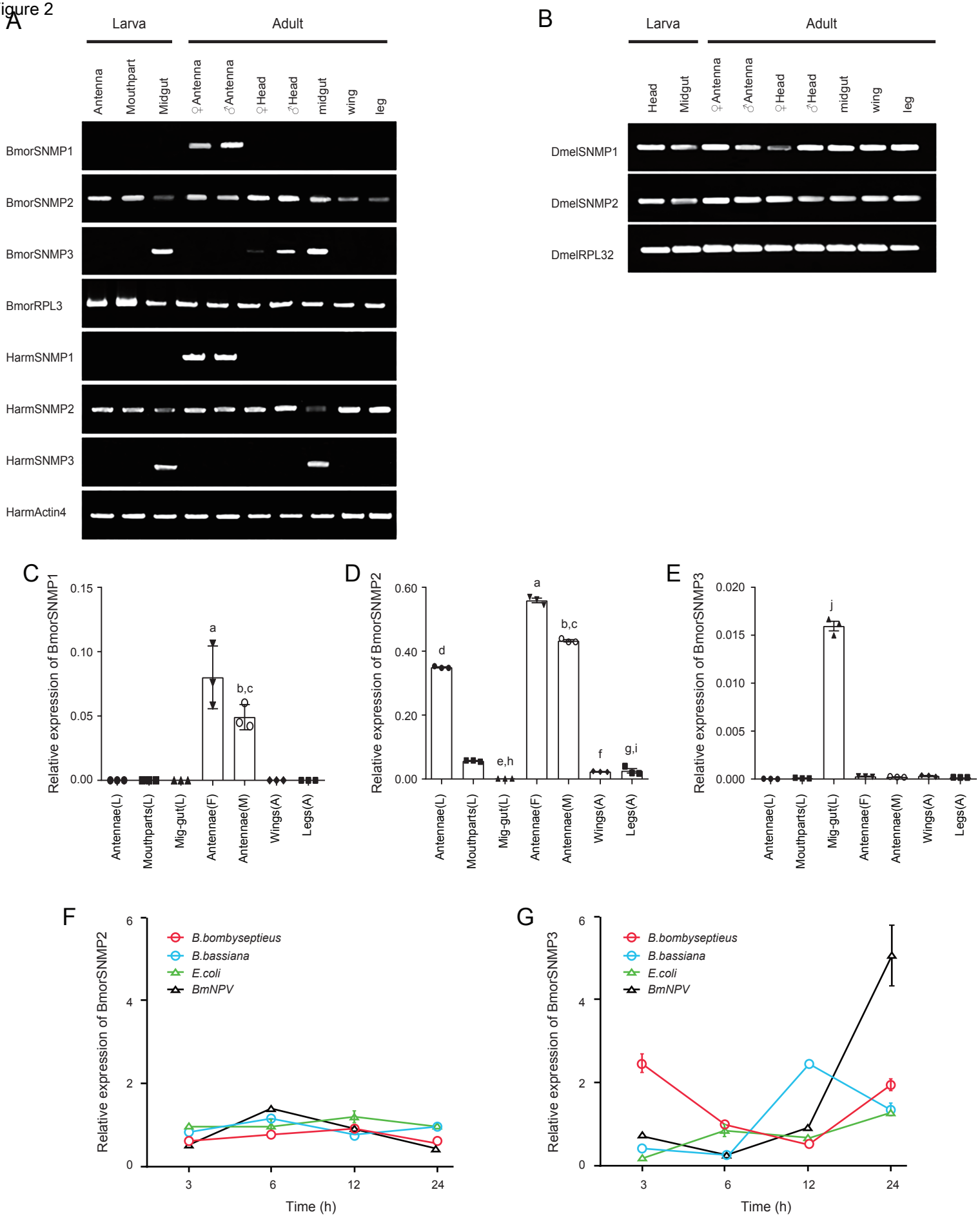

A
Larva
Adult
Antenna
Mouthpart
Midgut
♀Antenna
♂Antenna
♀Head
♂Head
midgut
wing
leg
BmorSNMP1
BmorSNMP2
BmorSNMP3
BmorRPL3
HarmSNMP1
HarmSNMP2
HarmSNMP3
HarmActin4
B
Larva
Adult
Head
Midgut
♀Antenna
♂Antenna
♀Head
♂Head
midgut
wing
leg
DmelSNMP1
DmelSNMP2
DmelRPL32
C
Relative expression of BmorSNMP1
0.15
0.10
0.05
0.00
a
b,c
Antennae(L)
Mouthparts(L)
Mig-gut(L)
Antennae(F)
Antennae(M)
Wings(A)
Legs(A)
D
Relative expression of BmorSNMP2
0.60
0.40
0.20
0.00
d
a
b,c
e,h
f
g,i
Antennae(L)
Mouthparts(L)
Mig-gut(L)
Antennae(F)
Antennae(M)
Wings(A)
Legs(A)
E
Relative expression of BmorSNMP3
0.020
0.015
0.010
0.005
0.000
j
Antennae(L)
Mouthparts(L)
Mig-gut(L)
Antennae(F)
Antennae(M)
Wings(A)
Legs(A)
F
Relative expression of BmorSNMP2
6
4
2
0
B.bombyseptieus
B.bassiana
E.coli
BmNPV
3
6
12
24
Time (h)
G
Relative expression of BmorSNMP3
6
4
2
0
B.bombyseptieus
B.bassiana
E.coli
BmNPV
3
6
12
24
Time (h)

Figure 5

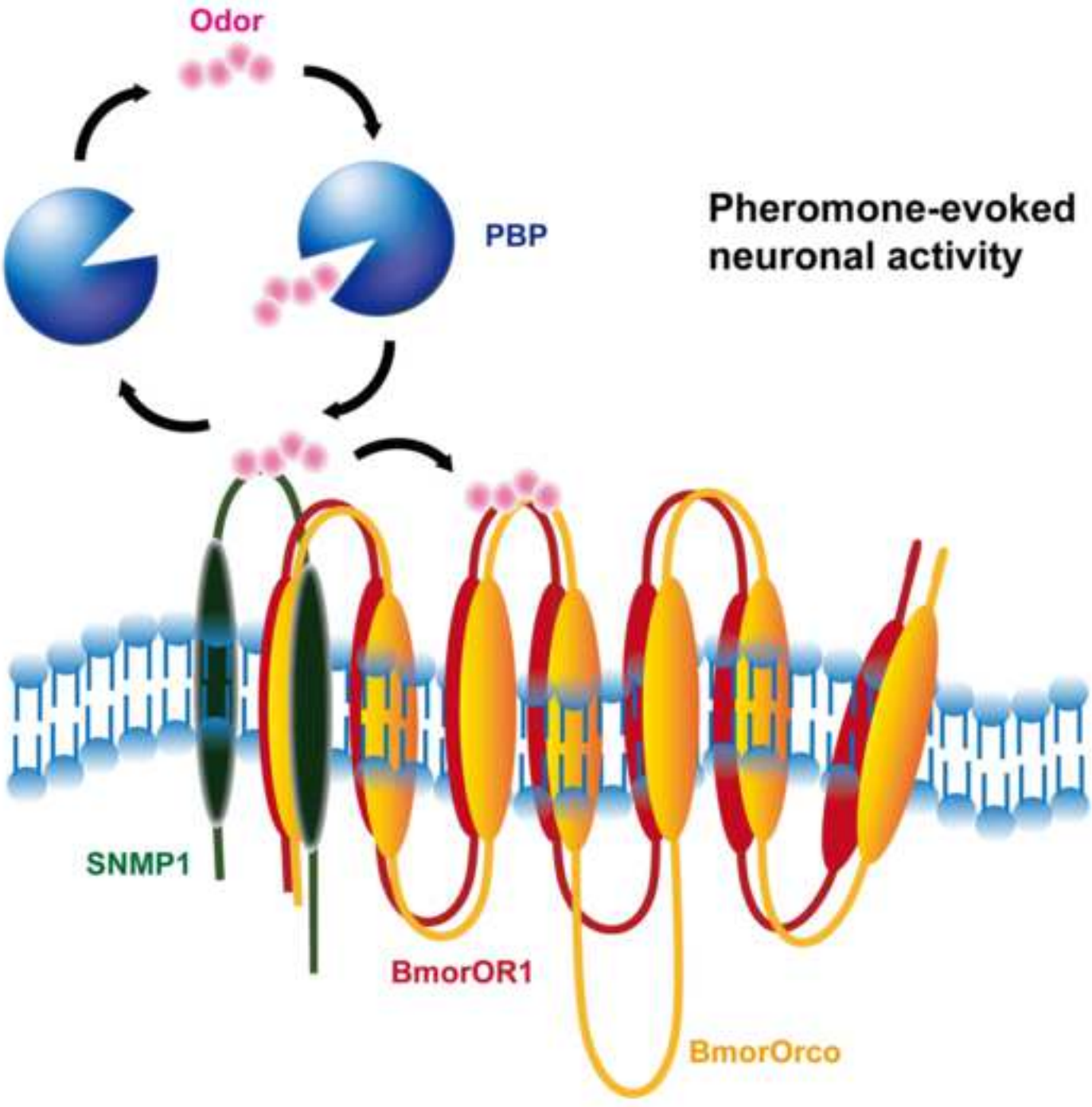


Figure 3

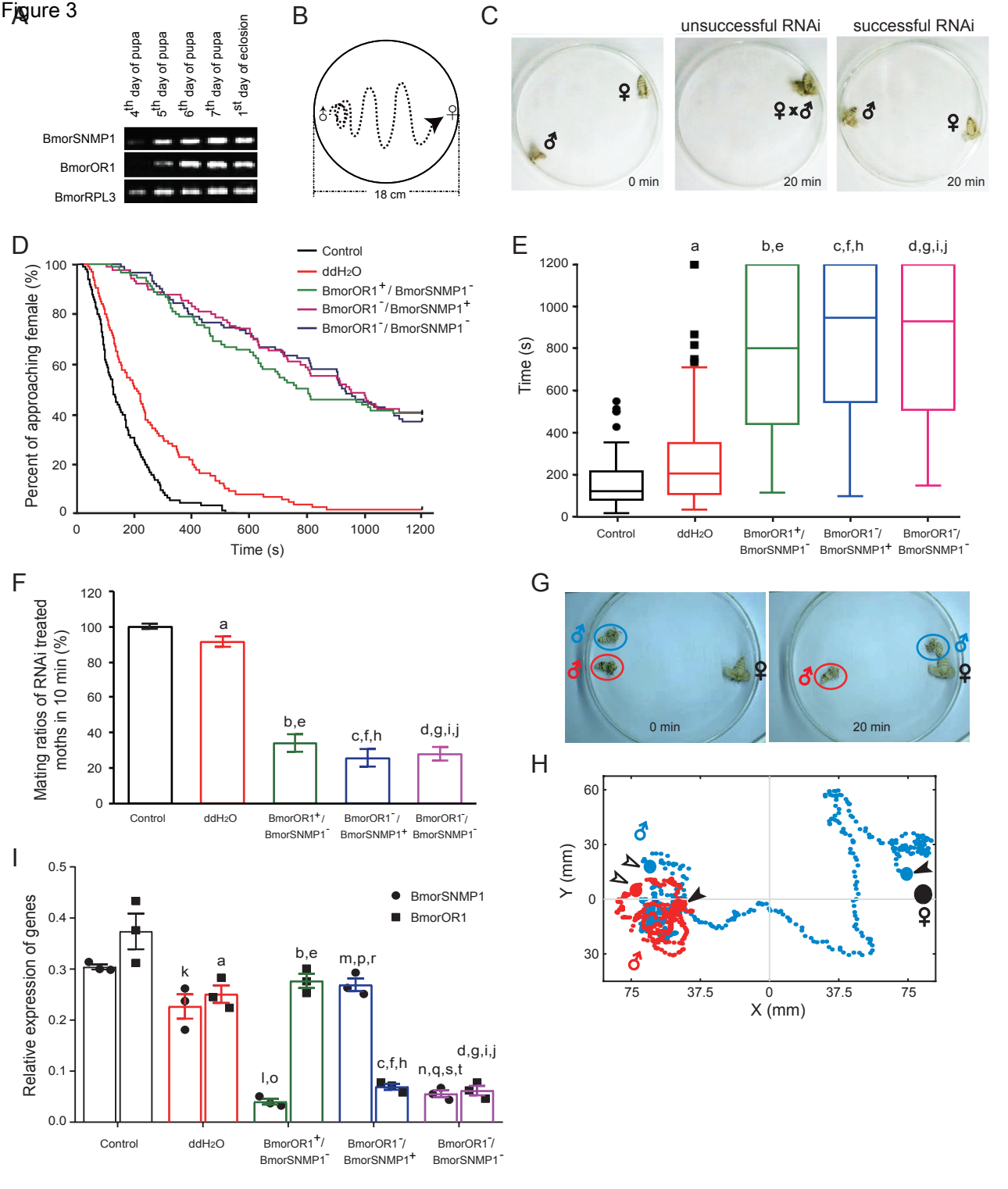
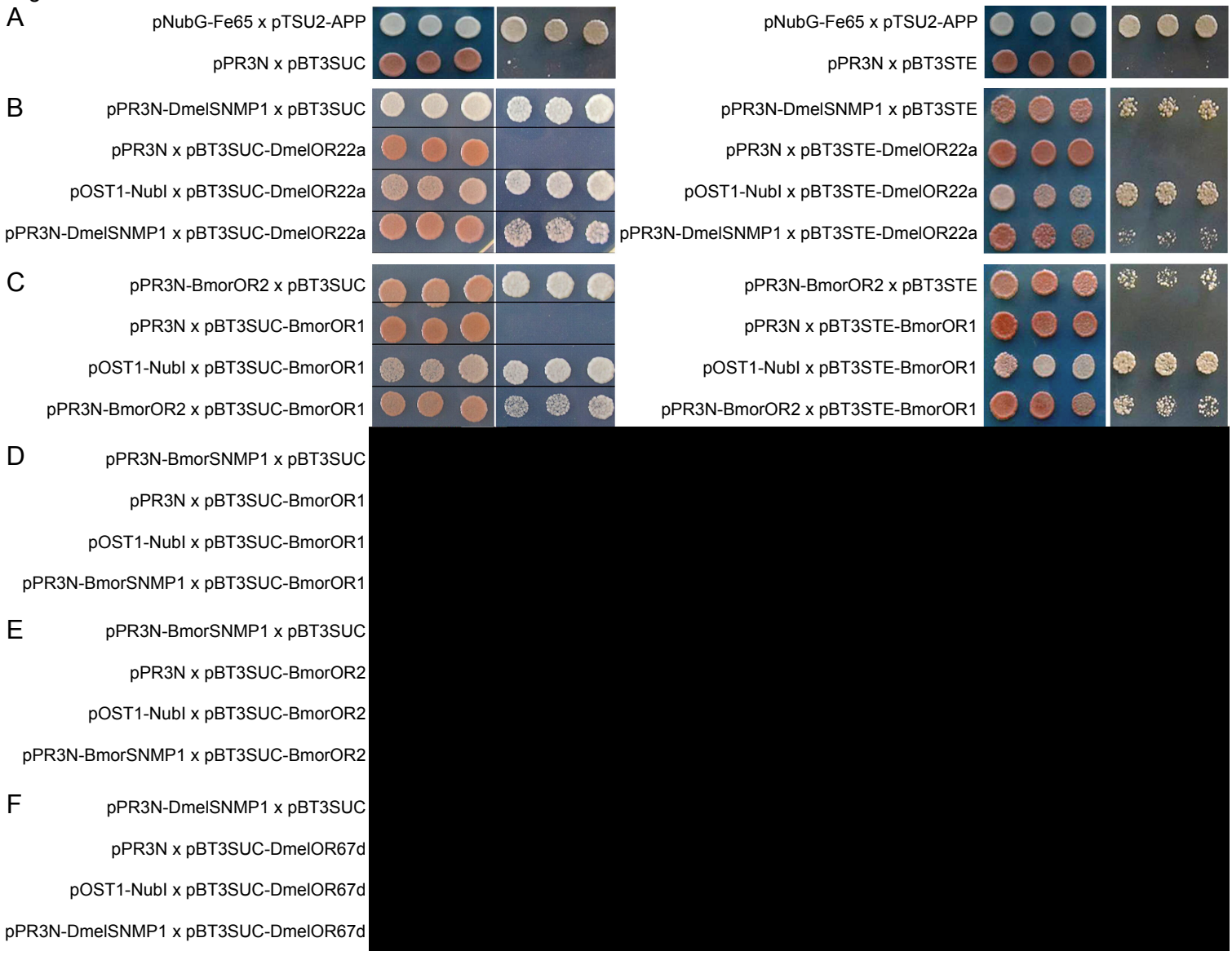

A
4th day of pupa
5th day of pupa
6th day of pupa
7th day of pupa
1st day of eclosion
BmorSNMP1
BmorOR1
BmorRPL3
B
18 cm
C
unsuccessful RNAi
successful RNAi
0 min
20 min
20 min
D
Control
ddH2O
BmorOR1+/ BmorSNMP1-
BmorOR1-/BmorSNMP1+
BmorOR1-/ BmorSNMP1-
Percent of approaching female (%)
Time (s)
E
Time (s)
a
b,e
c,f,h
d,g,i,j
Control
ddH2O
BmorOR1+/ BmorSNMP1-
BmorOR1-/ BmorSNMP1+
BmorOR1-/ BmorSNMP1-
F
Mating ratios of RNAi treated moths in 10 min (%)
a
b,e
c,f,h
d,g,i,j
Control
ddH2O
BmorOR1+/ BmorSNMP1-
BmorOR1-/ BmorSNMP1+
BmorOR1-/ BmorSNMP1-
G
0 min
20 min
H
Y (mm)
X (mm)
I
BmorSNMP1
BmorOR1
Relative expression of genes
k
a
b,e
m,p,r
l,o
c,f,h
n,q,s,t
d,g,i,j
Control
ddH2O
BmorOR1+/ BmorSNMP1-
BmorOR1-/ BmorSNMP1+
BmorOR1-/ BmorSNMP1-

Figure 4


SD-T/L
SD-T/L/H/A
SD-T/L
SD-T/L/H/A
A
pNubG-Fe65 x pTSU2-APP
pPR3N x pBT3SUC
pNubG-Fe65 x pTSU2-APP
pPR3N x pBT3STE
B
pPR3N-DmelSNMP1 x pBT3SUC
pPR3N x pBT3SUC-DmelOR22a
pOST1-NubI x pBT3SUC-DmelOR22a
pPR3N-DmelSNMP1 x pBT3SUC-DmelOR22a
pPR3N-DmelSNMP1 x pBT3STE
pPR3N x pBT3STE-DmelOR22a
pOST1-NubI x pBT3STE-DmelOR22a
pPR3N-DmelSNMP1 x pBT3STE-DmelOR22a
C
pPR3N-BmorOR2 x pBT3SUC
pPR3N x pBT3SUC-BmorOR1
pOST1-NubI x pBT3SUC-BmorOR1
pPR3N-BmorOR2 x pBT3SUC-BmorOR1
pPR3N-BmorOR2 x pBT3STE
pPR3N x pBT3STE-BmorOR1
pOST1-NubI x pBT3STE-BmorOR1
pPR3N-BmorOR2 x pBT3STE-BmorOR1
D
pPR3N-BmorSNMP1 x pBT3SUC
pPR3N x pBT3SUC-BmorOR1
pOST1-NubI x pBT3SUC-BmorOR1
pPR3N-BmorSNMP1 x pBT3SUC-BmorOR1
E
pPR3N-BmorSNMP1 x pBT3SUC
pPR3N x pBT3SUC-BmorOR2
pOST1-NubI x pBT3SUC-BmorOR2
pPR3N-BmorSNMP1 x pBT3SUC-BmorOR2
F
pPR3N-DmelSNMP1 x pBT3SUC
pPR3N x pBT3SUC-DmelOR67d
pOST1-NubI x pBT3SUC-DmelOR67d
pPR3N-DmelSNMP1 x pBT3SUC-DmelOR67d

Manuscript File

Supplementary Video 1

Click here to access/download

Video

S1 behavioural assays.mp4

Supplementary Video 2

Click here to access/download

Video

S2 RNAi.mp4

Supplementary Video 3

Click here to access/download

Video

S3 snmp single close to far.mp4

Supplementary Video 4

Click here to access/download

Video

S4 SNMP1 competition.mp4

Supplementary Video 5

Click here to access/download

Video

S5 BmOR1 competition.mp4